

-251-

Query: 63 D---NNSKIA 69  
           D   N+S IA  
 Sbjct: 62 DTTENDSLIA 71

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 581> which encodes the amino acid sequence <SEQ ID 582>. Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.0680 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below:

Identities = 21/61 (34%), Positives = 34/61 (55%)

Query: 1 MYNRLKELRKDKGLTQADLAKVINTNQSQYGYENGKTSLSIENSKILADFFGVSIPLYLL 60  
           MY R++ LR+D TQ +A +++ + + Y K E G+ +L + + VSI YLL  
 20 Sbjct: 1 MYPRI RNLRDNDFTQKFVANLLSFSHANYAKIERGEVALMADVLVQFYKLYNVSIDYLL 60  
  
 Query: 61 G 61  
           G  
 Sbjct: 61 G 61

25

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 174

- 30 A DNA sequence (GBSx0180) was identified in *S.agalactiae* <SEQ ID 583> which encodes the amino acid sequence <SEQ ID 584>. Analysis of this protein sequence reveals the following:

Possible site: 29  
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.5278 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 40 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 175

- 45 A DNA sequence (GBSx0181) was identified in *S.agalactiae* <SEQ ID 585> which encodes the amino acid sequence <SEQ ID 586>. Analysis of this protein sequence reveals the following:

Possible site: 60  
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.3762 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-252-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## 5 Example 176

A DNA sequence (GBSx0182) was identified in *S.agalactiae* <SEQ ID 587> which encodes the amino acid sequence <SEQ ID 588>. Analysis of this protein sequence reveals the following:

```

Possible site: 59
>>> Seems to have no N-terminal signal sequence
10    INTEGRAL    Likelihood = -9.66    Transmembrane    40 - 56 ( 33 - 65)
    INTEGRAL    Likelihood = -5.79    Transmembrane    62 - 78 ( 59 - 81)

----- Final Results -----
15          bacterial membrane --- Certainty=0.4864(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

20 A related GBS gene <SEQ ID 8505> and protein <SEQ ID 8506> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 7
McG: Discrim Score:      -16.96
GvH: Signal Score (-7.5): -2.95
Possible site: 57
25    >>> Seems to have no N-terminal signal sequence
ALOM program    count: 2 value: -9.66 threshold: 0.0
    INTEGRAL    Likelihood = -9.66    Transmembrane    33 - 49 ( 26 - 58)
    INTEGRAL    Likelihood = -5.79    Transmembrane    55 - 71 ( 52 - 74)
    PERIPHERAL  Likelihood = 10.87      14
30    modified ALOM score: 2.43

*** Reasoning Step: 3

----- Final Results -----
35          bacterial membrane --- Certainty=0.4864(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 40 Example 177

A DNA sequence (GBSx0183) was identified in *S.agalactiae* <SEQ ID 589> which encodes the amino acid sequence <SEQ ID 590>. Analysis of this protein sequence reveals the following:

```

Possible site: 31
45    >>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.3276(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50

```

The protein has no significant homology with any sequences in the GENPEPT database.

-253-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 178

- 5 A DNA sequence (GBSx0184) was identified in *S.agalactiae* <SEQ ID 591> which encodes the amino acid sequence <SEQ ID 592>. Analysis of this protein sequence reveals the following:

Possible site: 44  
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.3482 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 A related GBS nucleic acid sequence <SEQ ID 9509> which encodes amino acid sequence <SEQ ID 9510> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA30291 GB:X07371 RepM protein (AA 1 - 314) [Staphylococcus aureus]  
20 Identities = 89/283 (31%), Positives = 145/283 (50%), Gaps = 26/283 (9%)  
  
Query: 67 KVSLDNITMTAYIKSKKYLAMKQLIETHLAITVQTAMTDMFRATTGDIHVVLHMNYDKQ 126  
K+S D +T+ + + + I + + F+A + +++ YDK  
Sbjct: 42 KLSFDAMTIVGNLNKNSAKKLSDFMSLDPQIRLWDILQTKFKAKA---LQEKVYIEYDKV 98  
  
25 Query: 127 KGQDRKARPFRLFPNPNKLRLVDSEII---DTIIPFLEDISISRADLAFDLFEVDCSEF- 182  
K R R+EFNPNKL E++ II ++ED +R DLAFD FE D S++  
Sbjct: 99 KADTWDRRNMRFNPNKL--THDEMLWLKHNIIDYMEDDGFTRLDLAFD-FEDDLSDYY 155  
  
30 Query: 183 -VLEKKGRPTATKEFRSSTGTLET KYLGAPRSEKQVRLYNKKKEQLQNGTDKDKDFASQF 241  
+ EK + T F +TG ETKY G+ S + +R+YNKKKE+ +N D D +++  
Sbjct: 156 ALSEKALKRTV---FFGTGKAETKYFGSRDSNRFIRIYNKKKERKENA---DVDVSAE- 208  
  
35 Query: 242 KHWWRLEFQLRSRIDEIFEVI-DTIIFKP--FNLKGLSIETQIYLTALIHDKNIWKKLH 298  
H WR+E +L+ +D D I KP L+ L + +YL L+H+++ W +LH  
Sbjct: 209 -HLWRVEIELKRDMDVYWNCFNDLHILKPAAWATLESKEQAMVYL--LLHEESKWGELH 265  
  
Query: 299 RNTRARYKKILETHQTSDDYLGLLKDLLKHERPRLNQLAYY 341  
RN+R +YK+I++ + S D L+K L L+ Q+ ++  
40 Sbjct: 266 RNSRRKYKQIIQ--EISSIDLTLMKSTLTDNEENLQKQINFW 306

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 45 Example 179

A DNA sequence (GBSx0185) was identified in *S.agalactiae* <SEQ ID 593> which encodes the amino acid sequence <SEQ ID 594>. Analysis of this protein sequence reveals the following:

Possible site: 32  
>>> Seems to have no N-terminal signal sequence

50 INTEGRAL Likelihood = -15.55 Transmembrane 137 - 153 ( 133 - 157)

----- Final Results -----  
bacterial membrane --- Certainty=0.7220 (Affirmative) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8507> and protein <SEQ ID 8508> were also identified. Analysis of this protein sequence reveals the following:

\*\*\* Reasoning Step: 3

----- Final Results -----

```

bacterial membrane --- Certainty=0.7220(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

ORF01844(292 - 702 of 1074)
EGAD|124517|132830(149 - 295 of 435) apolipoprotein A-IV {Mus musculus}
GP|191889|gb|AAA37216.1|M64250 apolipoprotein A-IV {Mus musculus castaneus}
%Match = 4.6
%Identity = 30.0 %Similarity = 54.6
Matches = 39 Mismatches = 53 Conservative Sub.s = 32

```

NSSNIRY\*LFRFAERLVEA\*KTKTRKSARLLWG\*DRQK\*LSTLLLKIQQYQGVTRSKMRIKD YADSLGVSSQSIIYKRIRS  
| : | : | : | : | : | :  
LRDRMMPHANKVTQTFGENMQKLQEHLKP YA VDLQDQINTQTQEMKQLTPYIQRMQTTIKENVDNLHTSMMLPLATNLKD  
120            130            140            150            160            170            180

[illegible]

615            645            672            702            732            762            792            822  
 ---RIDRLADKLTPLEDNQNLVQKNYE-LLNYVRSLERQKLLLIIALAVMVTLLVAIWLAIFF\*WQLSDNAKRPTKGTA  
 :||:|    |:||:|    |:|:|    |    :|||    :|||  
 VSAKIDQLQKNLAPLVEDVQSKVGNTTEGLQSKLDKLNRLQLEQQVEEFRRTVEPMGEMFNKALVQQLEQFRQQLGPNSGE  
 270            280            290            300            310            320            330

SEQ ID 8508 (GBS405) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 171 (lane 4; MW 46kDa – 2 bands) and in Figure 177 (lane 7; MW 46kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 76 (lane 5; MW 21kDa).

GBS405-GST was purified as shown in Figure 218, lane 8.



-255-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 180

A DNA sequence (GBSx0186) was identified in *S.agalactiae* <SEQ ID 595> which encodes the amino acid sequence <SEQ ID 596>. Analysis of this protein sequence reveals the following:

Possible site: 18  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10                   bacterial cytoplasm --- Certainty=0.3406(Affirmative) < succ>  
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15           >GP:CAA33713 GB:X15669 pre protein (AA 1-494) [Streptococcus  
                  agalactiae]  
          Identities = 171/402 (42%), Positives = 250/402 (61%), Gaps = 46/402 (11%)

20           Query: 1   MSYVVARMAKYKSGQLTAIYNHNERIFKNHSNKEIDVEKSHLNYELTNRDQAQNYHKQIK 60  
                  MSY+VARM K K+G L   + HNER+F+ HSNK+I+   +SHLNYELT+RD++ +Y KQIK  
          Sbjct: 1   MSYMVARMQMKAGNLGGAFKHNERVFETHSNKDINPSRSHLNYELTDRDRSVSYEKQIK 60

25           Query: 61   EHINENRLSTRGVRKDAILCNEWIITSDKTFDFSLDEKQTRFEFFETAKDYFAEKYGDANI 120  
                  +++NEN++S R +RKDA+LC+EWIITSDK FF+ LDE+QTR FFETAK+YFAE YG++NI  
          Sbjct: 61   DYVNENKVSNNRAIRKDAVLCDEWIITSDKDFEKLDEEQTRTFETAKNYFAENYGESNI 120

30           Query: 121   AYARVHLDSTPHMHLGIVPMKNGKLSSKALFGNKEKLVAIQDELPKYLNEHGFNLQGE 180  
                  AYA VHLDESTPHMH+G+VP +NGKLSSKA+F ++E+L   IQ++LP+Y+++HGF L+RG+  
          Sbjct: 121   AYASVHLDSTPHMHMGVVPFENGKLSSKAMF-DREELKHIQEDLPRYMSDHGFELERGK 179

35           Query: 181   IGSKKKHLETAEFKEKQRLLDNADRKLADKHEELKALDDKISNV-NDTIA----- 229  
                  + S+ KH   AEFK   ++   +L +K+   +D++   + NDT A  
          Sbjct: 180   LNSEAKHKTVAEFKRAMADME-LKEELLEKYHAPPFVDERTGELNNDTEAFWHEKEFADM 238

40           Query: 230   -DKESRLKEL---EAKEWDVAGDLKQYELEKQSLAESIEDIKDIELLQLDRIQKEDLVKQ 285  
                  + +S ++E   E +W   KQY+ E + L S + ++D   D   E+L+ +  
          Sbjct: 239   FEVQSPIRETTNQEKMDWLR----KQYQEELKKLESSKKPLED-----DLSHLEELLDK 288

45           Query: 286   SFDGKLKMDKETYNRLFQTASKHASSNAELKRDLVKAQSQNNHLSRELLNHRKTAENIK 345  
                  +K+D E           AS+ AS           +L KA+   N L   NH K+ E I+  
          Sbjct: 289   KTKEYIKIDSE-----ASERAS-----ELSKAEGYINTLE----NHKSLEAKIE 329

          Query: 346   LSQENRKLKDKVKMLDEQVKILNKSLSVWKEKAKEFMPKQVY 387  
                  + +   +K K   + K LN+S   + K F+ K+ Y  
          Sbjct: 330   CLESNDNLQLEKQKATKLEAKALNESELRELKPKKNFLGKEHY 371

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 597> which encodes the amino acid sequence <SEQ ID 598>. Analysis of this protein sequence reveals the following:

LPXTG motif: 2025-2030

Possible site: 52

>>> Seems to have no N-terminal signal sequence

INTEGRAL   Likelihood = -10.08   Transmembrane 2034 -2050 (2030 -2053)  
INTEGRAL   Likelihood = -6.05   Transmembrane 21 - 37 ( 20 - 39)

----- Final Results -----

60                   bacterial membrane --- Certainty=0.5034(Affirmative) < succ>  
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
                  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAD03320 GB:AF067776 extracellular matrix binding protein  
[Abiotrophia defectiva]

Identities = 362/1396 (25%), Positives = 591/1396 (41%), Gaps = 87/1396 (6%)

Query: 636 KAEVKLKEAHEATKQAIEKDPWLSPEQKKAQKEKAKARLDEGLKALKKAADSLEILKVTEE 695  
+A+ + A +A AI+ + L+ E+K A+K K +A + L + A K T

Sbjct: 636 EAKNAVNNAAKAKNTAIDNNNNLTAEKAAEKAKVEAAKNATLAGIDQA-----KTAA 689

Query: 696 AFVDKEKNPDSIPNQHKAGTADQARKQALDSDLKEVQKELESIDNDNTLTDEKAAAKKK 755  
+ K I + A A AL+ + ++ I LT +EK A +

Sbjct: 690 RNAAQNKGTTDINAVNFPVPAKPAANAAL- -QA AVNKINEISQRPDLTREETQAFMDQ 746

Query: 756 VNDAVDVAKQTAMEANSYEDLTITIKDEFLS---NLPHKQGTPLKDDQSDAIAELEKKQBE 812  
V A D A A + + +T+ +D+ L+ NLP TP + +A+ + +

Sbjct: 747 VRTARDAAMAKVASAANNQAVTSARDQGLNAVNNLP---TPAA-KYPEALGHVRQAADA 801

Query: 813 IEKAIIEGDKTLPRDEKEKQIADSKERLKSQTKVKDAKNADAIKKAFEEGKVNIPQAHIP 872  
+AI + L +E+ + + + + KA +G I

Sbjct: 802 KRQAIRDNANLTAEEQADALRQVDAQAQTAEEAIAINQHTNATLAKADSDGVKAI----- 855

Query: 873 GDILN---KDKEKLLAELKQKADDTFEKAIDVDKTLTDEKKEQKVKTKAELEKAKTDVKNT 929  
D+N + K L+Q A +AI+ + LT++EK + + L AKT V+

Sbjct: 856 NDINPQPRSKPAANQALEQVAAAKROAINNNQLTDEEKAQAIQQVDQALANAKTQVQAA 915

Query: 930 QTREELDKKVPKKAIEDTHVKNLEGVKNKAIEDLKAHTETVAKINGDDTLDKATKE 989  
+++ AI + + +G K +AI ++ A ++ G + L +

Sbjct: 916 NDNNGVNVQAKTAGTTAINNINPQGTQ---KAQAIAAIEAAEQAKRLELQGRNDLTTEERN 972

Query: 990 AQVKREADKALAAGKDAITKADDADKVVSTAVTEHTPKIKAAHKTGDLKKAQVDANTALDKA 1049  
+ + A KDA+ +A + V+ A +I+ + T +K DA A+D+A

Sbjct: 973 NALADLTAKAQAAKDAVNQARNNTGVAGAKDNGVAQIQGINPTAVVKP---DARNAIDQA 1029

Query: 1050 AEKERGEINKDATLTTEDEKAKQLKEVETALTAKDNVKAAKTADAINDARDKGVATIDAV 1109  
A + E + LT E+KA +K+V+ A AK + A + +N+A ++G A I A+

Sbjct: 1030 ARDKAEFQANTKLTDEEKAAAIAKVVQDAARDAKAAIDRAGSNGDVNNAVNQKAAIQAI 1089

Query: 1110 HKAGQDLGARKSGQVAKLEAAKATKDKISADPTLTSKEKEEQSKAVDAELKKAIEAVNA 1169  
+ K A ++ AA A K I+A+ LT +EK K V+ E KA AV+A

Sbjct: 1090 KALDDSQPSAKDTAKAAIQNAADAKKAITANNALTQEEKAAAIAKQVEDEAAKQAAVDA 1149

Query: 1170 ADTADKVDDALGEGVTDIKNQHKSGDSIDARREAHGKELDRVAQETKGAIEKDPPTLTTEE 1229  
+ + VD A +G+ I + ++ + +D+ A + K I D TLT EE

Sbjct: 1150 SRSKADVDRAKDQGLQKISDV---PAVQPPKLNAIAAVDQAATDKKAVINNDTTLTQEE 1205

Query: 1230 KAKQVKDVDAKERGMKLNKADADALDKAYGEGVTDIKNQHKSGDPVDARRGLHNKSI 1289  
K ++ VD + +N+A + +G I N ++ A + ++

Sbjct: 1206 KEAAIRKVDEEAAKARQAINDATSNADVAKQAGTQAINNVPT---PAKNAAKAAV 1261

Query: 1290 DEVAQATKDAITADTTLTEAEKETQRGNVDKREATKAKEELAKAKDADALDKAYGDGVTSI 1349  
++ A A K AI D LT EK+ VD+E KA++ + A + +G +I

Sbjct: 1262 EQAADAKQAIENDPNLTROEKDAAIAKVDQETNKARQAI DAATTNADVTAKQNEGTQAI 1321

Query: 1350 KNQHKSGKGLDVRKDEHKKALEAVAKRVTAIEADPTLTPEVREQQAEVQKELELATDK 1409  
++ K K + K A+ A+ + IE DP LT E ++ KA+V E A +

Sbjct: 1322 NAVPQTPKA---KTDAKNAVTAQAEDKKSAINENDPNLTREETDAKAKVDAEATKAKNA 1377

Query: 1410 IAEAKDADEADKAYGDGVTAIENAHVIGKIEARKDLAKKDLAEAAAKTKALIIEDKTLT 1469  
I A D+ +G AI + + + +A+ D AK + +AA + K I D LT

Sbjct: 1378 IDAATSNDDETAKQNEGTQAI---NAVPTPKAKTD-AKNVTAQADRKKDAIENDPNLT 1433

Query: 1470 DDQRKEQLLGVDTVEYAKGIENIDAADKAAAGVDKAYSVDGVRDILAQYKEGQNLNDRRNAK 1529  
+++ VD E K + IDAA A V ++G + I + + AK

Sbjct: 1434 REEKVAAKAKVDAAEAKKAKDAIDAATSNAVDVTAKQNEGTKAI---NDVPQTPAKTDAK 1489

Query: 1530 EFLLEADKVTKLINDDPTLTTHDQKVDQINKVEQAKLDAIKSVDDAQTADAINDALGKGI 1589

-257-

+ + AD I DP LT ++K KV+ A ++D A + + +G  
 Sbjct: 1490 NAVTQAADAKKDAIEKDPNLTREEKDAAKAKVDAAEAKKAKDAIDAATSNADVTAQNENGT 1549  
 Query: 1590 ENINNQYQHGDGVDVRKATAKGDLEKEAAKVKALIAKDPTLTQADKDKQTAAVDAAKNTA 1649  
 + IN+ Q K AK + + A K I KDP LT+ +KD A VDA A  
 Sbjct: 1550 KAINDVDPQ---TPTAKTDAKNAVTAQADAKKDAIEKDPNLTREEKDAAKAKVDAAEAKKA 1605  
 Query: 1650 IAAVDKATTTEGINQELGKGITAINKAYRPGEGVKARKEAAKADLEKEAAKVKALITNDP 1709  
 A+D AT+ + + G AIN + K AK + + A K I ND  
 Sbjct: 1606 KDAIDAATSNADVTAQKDAGKNAINAVPQ---TPTAKTDAKNAVTAQADAKKDAIENDA 1661  
 Query: 1710 TLTAKDK-AKQTEAVAKALKAAIAAVDKATTAEINQELGKGITAINKAYRPGEGVKARK 1768  
 LT+ +K A + + A+A KA A+D AT+ + + +G AIN + K  
 Sbjct: 1662 NLTREEKDAAKAKVDAAEATKAK-NAIDAATSNADVTAQNENGTKAINDVDPQ---TPTAK 1716  
 Query: 1769 EAAKADLEREAARKVREAIANDPTLTAKDK-AKQTEAVAKALKAAIAAVDKATTAEINQE 1827  
 AK +++ A + AI NDP LT+ +K A + + A+A KA A+D AT+ + +  
 Sbjct: 1717 TDAKNAVDAQATDKKSAIENDPALTREEKDAAKAKVDAAEATKAK-NAIDAATSNADVTAQ 1775  
 Query: 1828 LGKGITAINKAYRPGEGVEAHKEAAKANLEKVKAKETKALISGDRYLSETEKAVQKQAVEQ 1887  
 G AIN + K AK +++ A + KA I D L+ EK K V+  
 Sbjct: 1776 KDAGKNAINAVPQ---TPTAKTDAKNAVDAQATDKKAAIENDPALTREEKDAAKAKVDA 1831  
 Query: 1888 ALAKALGQVEAAKTVEAVKLAENLGTVAIRSAYVAGLAKDTTQATAALNEAKQAAIEALK 1947  
 KA ++AA + V ++ G KD A AK A A+  
 Sbjct: 1832 EAKKAKDAIDAATSNADVTAQKDAG-----KDAINAVPQTPTAKTDAKNAV 1878  
 Query: 1948 QAAAEETLAKITTDALTEAQAEQSENVLALKTAIATVRSQAQSIASVKEAKDKGITAIR 2007  
 QAA + + I D LT +K V KA + +A S A V + +G AI  
 Sbjct: 1879 QAATDKKSAIENDPALTREEKDAVAKAKVDAAEAKKAKDAIDAATSNADVTAQTEGTQAIN 1938  
 Query: 2008 AAYVPNKAVAKSSSAN 2023  
 A VP AK+ + N  
 Sbjct: 1939 A--VPQTPTAKTDAKN 1952

An alignment of the GAS and GBS proteins is shown below:

Identities = 77/396 (19%), Positives = 157/396 (39%), Gaps = 48/396 (12%)

Query: 42 LNYELTNRDQAQNYHKQIKEHINENRLSTRGVRKDAILECNEWIITSCKTFFDSLDEKQTR 101  
 L++E+ + ++QN K+I + + D E +I K +++ EK T  
 Sbjct: 338 LDFEILH-PRSQNVSKKISKQVEAKPF-----DPASYKEKVIAKLKPVEATSEKITN 389  
 Query: 102 EFF--ETAKDYFAEKYGDANIAYARVHLDESTPHMHLGIVPMKNGKLSSKALFG--NKEK 157  
 + + E AKD +K + I+ G V + +A+ NK  
 Sbjct: 390 DAWLDENAKDLQKQKLEEQYIS-----GKVAISEAGTKQEAIDAAYNKYS 434  
 Query: 158 LVAIQDELPKYLNEHGFNLQGEIGSKKKHLETAEFKEKQRLLDN---ADRKLADKHEEL 214  
 D LP + N + + ++ ++T + K D K K E L  
 Sbjct: 435 SQTPDPSLPSQYKQG--NKENEQEKGQRDLIQTRDLTLKAIQEDKWLTEQEKTIQKEEAL 492  
 Query: 215 KALDDKISNVNDTIADKESRLKELEAKEDAVGDLKQYE-----LEKQSLAESIE 264  
 KA + I +VN T++ ++ + + + K + + K+Y EK+ A E  
 Sbjct: 493 KAFETGIESVNTQVSLEQLKQRLIVYKASEKDSEKKEYPESIPNQHIPGKEKEVKAQKE 552  
 Query: 265 DIKDIELLQLDRIQKEDLVQKSFQDGLKMDKETYNRLFQTASKHASSNAELKRDVLVKAQS 324  
 ++K + L++I ++ + E + Q A K A + +L+ DL S  
 Sbjct: 553 ELKKLHDTTLEKINQDKWLTPDQQAQQLKQAEVTFKKGQEAIKSAQTTLTQLETDLADYVS 612  
 Query: 325 QNNHLSRELLNHRKTAENIKLSQENRKLKDKVKMLDEQVK----ILNKSLSVWKEKAKE 380  
 +N + + K+ K+ +++ KKK+ + + ++ + + KEKAK  
 Sbjct: 613 ENEGKGNIPDKYKSGNKDDLNVNKAIEVKKLEAHEATKQAEKDPWLSPEQKKAQKEKAKA 672  
 Query: 381 FMPKQVYRETLIIINTLNPIGLAKTAIRQVKMVD 416  
 + + + + L ++L + + + A +K DS  
 Sbjct: 673 RLDEGL--KALKAADSLEILKVTEAFVDKEKNPDS 706

-258-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 181

A DNA sequence (GBSx0187) was identified in *S.agalactiae* <SEQ ID 599> which encodes the amino acid sequence <SEQ ID 600>. Analysis of this protein sequence reveals the following:

Possible site: 42  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10                   bacterial cytoplasm --- Certainty=0.2544 (Affirmative) < succ>  
                    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 182

A DNA sequence (GBSx0188) was identified in *S.agalactiae* <SEQ ID 601> which encodes the amino acid sequence <SEQ ID 602>. Analysis of this protein sequence reveals the following:

Possible site: 57  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25                   bacterial cytoplasm --- Certainty=0.2045 (Affirmative) < succ>  
                    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 603> which encodes the amino acid sequence <SEQ ID 604>. Analysis of this protein sequence reveals the following:

Possible site: 57  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35                   bacterial cytoplasm --- Certainty=0.2045 (Affirmative) < succ>  
                    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

40 An alignment of the GAS and GBS proteins is shown below:

Identities = 102/111 (91%), Positives = 107/111 (95%)

Query: 1 MDYKKYQIIYAPDVLEKLKEIRDYISQNYSSSTSGQHMEQIISDIEKLEVFPEVGFDAD 60

+DYKKYQIIYAPDVLEKLKEIRDYISQNYSSSTSGQ KMEQIISDIEKLEVFPEVGFDAD

45 Sbjct: 1 LDYKKYQIIYAPDVLEKLKEIRDYISQNYSSSTSGQRKMEQIISDIEKLEVFPEVGFDAD 60

Query: 61 KYGSKISKYHSTRGYTLSKDYIVLYHIEEENRVVIDYLLPTRSDYMKLFK 111

KYGSKI YHST+GYTLSKDYIVLYHIE EENR+VIDYLLPT+SDY+KLFK

Sbjct: 61 KYGSKIIHYHSTKGYTLSKDYIVLYHIEEENRIVIDYLLPTQSDYIKLFK 111

50

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 183

A DNA sequence (GBSx0189) was identified in *S.agalactiae* <SEQ ID 605> which encodes the amino acid sequence <SEQ ID 606>. Analysis of this protein sequence reveals the following:

Possible site: 13  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1621(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 607> which encodes the amino acid sequence <SEQ ID 608>. Analysis of this protein sequence reveals the following:

Possible site: 22  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1596(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 91/95 (95%), Positives = 93/95 (97%)

Query: 1 MVTAEKNRAVTFQANKELVSEAMTVLNKKNLTLSSALRLFLQNVVVTNEVDLLTEEELEK 60  
M T +KNRAVTFQANKELVSEAMTVLNKKNLTLSSALRLFLQNVVVTNEVDLLTEEELEK

Sbjct: 1 MTTVKKNRAVTFQANKELVSEAMTVLNKKNLTLSSALRLFLQNVVVTNEVDLLTEEELEK 60

Query: 61 EKLFKQFQAEINKNIEDVRQGFYTSEEVRSELGL 95  
EKLFKQFQAEINKNIEDVRQGFYTSEEVR+ELGL

Sbjct: 61 EKLFKQFQAEINKNIEDVRQGFYTSEEVRSELGL 95

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 184

A DNA sequence (GBSx0190) was identified in *S.agalactiae* <SEQ ID 609> which encodes the amino acid sequence <SEQ ID 610>. Analysis of this protein sequence reveals the following:

Possible site: 56  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4568(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9513> which encodes amino acid sequence <SEQ ID 9514> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

-260-

>GP:CAA46375 GB:X65276 ORFA1 [Clostridium acetobutylicum]  
Identities = 36/91 (39%), Positives = 51/91 (55%)

Query: 2 MSQIKLTPEELRISAQKYTTGSQSITDVLTVLTQEQAVIDENWDGTAFDSFEAQFNELSP 61  
M+QI +TPEEL+ AQ Y + I + + + I E W G AF ++ Q+N+L  
Sbjct: 1 MAQISVTPPEELKSQAQVYIQSKEEIDQAIQKVNMSMNSTIAEEWKGAQAYLEQYNQLHQ 60

Query: 62 KITQFAQLLEDINQQLLKVADVVEQTDSIDIA 92  
+ QF LLE +NQQL K AD V + D+ A

Sbjct: 61 TVVQFENLLESVNQQLNKYADTVAERDAQDA 91

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 185

A DNA sequence (GBSx0191) was identified in *S.agalactiae* <SEQ ID 611> which encodes the amino acid sequence <SEQ ID 612>. Analysis of this protein sequence reveals the following:

Possible site: 21  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.4523(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 186

A DNA sequence (GBSx0192) was identified in *S.agalactiae* <SEQ ID 613> which encodes the amino acid sequence <SEQ ID 614>. Analysis of this protein sequence reveals the following:

Possible site: 44  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.5339(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 187**

A DNA sequence (GBSx0193) was identified in *S.agalactiae* <SEQ ID 615> which encodes the amino acid sequence <SEQ ID 616>. This protein is predicted to be chromosome assembly protein. Analysis of this protein sequence reveals the following:

```

5   Possible site: 61
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.4620(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 188**

A DNA sequence (GBSx0194) was identified in *S.agalactiae* <SEQ ID 617> which encodes the amino acid sequence <SEQ ID 618>. Analysis of this protein sequence reveals the following:

```

20   Possible site: 46
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
25      bacterial cytoplasm --- Certainty=0.4511(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 189**

A DNA sequence (GBSx0195) was identified in *S.agalactiae* <SEQ ID 619> which encodes the amino acid sequence <SEQ ID 620>. Analysis of this protein sequence reveals the following:

```

35   Possible site: 20
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
40      bacterial cytoplasm --- Certainty=0.5249(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 190**

A DNA sequence (GBSx0196) was identified in *S.agalactiae* <SEQ ID 621> which encodes the amino acid sequence <SEQ ID 622>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 14
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.3542(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9515> which encodes amino acid sequence <SEQ ID 9516> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 191**

20 A DNA sequence (GBSx0197) was identified in *S.agalactiae* <SEQ ID 623> which encodes the amino acid sequence <SEQ ID 624>. Analysis of this protein sequence reveals the following:

```

   Possible site: 15
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
25          bacterial cytoplasm --- Certainty=0.3098(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

30 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 192**

35 A DNA sequence (GBSx0198) was identified in *S.agalactiae* <SEQ ID 625> which encodes the amino acid sequence <SEQ ID 626>. This protein is predicted to be rgg protein. Analysis of this protein sequence reveals the following:

```

   Possible site: 59
   >>> Seems to have no N-terminal signal sequence

40   ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.3177(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

45 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAA26968 GB:M89776 rgg [Streptococcus gordonii]
Identities = 74/277 (26%), Positives = 142/277 (50%)

```



-263-

5  
 Query: 7 IFREFRLNRQFSLKQVASNELSVS QLSRFRGESDLSLTKFLGALEAIDL SISEFMDRVN 66  
 I + R ++ SLK+VA+ ++SV+QLSR+ERG S L++ F L + +S++EF +  
 Sbjct: 10 ILKI IRESKNMSLKEVAAGDISVAQLSR YERGISSLTVDSFYSC LRNMSVSLAEFQYVYH 69

10  
 Query: 67 KYQKSDQISLMSQMAQYHYQRDVAGLEKMISVEEGKLKDDSSDIRCRLNIVLFRGMICEC 126  
 Y+++D + L ++++ + ++ LE +++ E ++ +LN ++ R + C  
 Sbjct: 70 NYREADDVVLSQKLSEAQRENNIVKLESILAGSEAMAQEFPEKKNYKLNTIVIRATLTSC 129

15  
 Query: 127 DSSRKMSEEDLCFLSDYLFQKDSWEISDYILIGNLYRYYNTRHICQLVKEVINQKEYYRD 186  
 + ++S+ D+ FL+DYLF + W + L N + E+IN+ ++Y +  
 Sbjct: 130 NPDYQVSKGDI EFLTDYLF SVEWGRYELWLF TNSVNL LTLETLETFASEMINRTQFYNN 189

20  
 Query: 187 IYTNRNVEATLLNVVETLIERRALEEATFFLEKVEALLNNERNAYHRIILLYEKGFLAY 246  
 + NR + LLNVV IE L+ A FL ++ E + Y R+++ Y K +Y  
 Sbjct: 190 LPENRRRIIKMLLNVSACIENNLQVAMKFLNYIDNTKIPETDLYDRVLIKYHKALYSY 249

25  
 Query: 247 AKGDSRGIQSMKQAIFCFQAIGSKHHVENFQEHFNRV 283  
 G+ ++Q + F+ + S +E F R+  
 Sbjct: 250 KVGNP HARHDIEQCLSTFEYLD SFGVARKLKEQFERI 286

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 627> which encodes the amino acid sequence <SEQ ID 628>. Analysis of this protein sequence reveals the following:

25  
 Possible site: 29  
 >>> Seems to have no N-terminal signal sequence

30  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3792(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 79/275 (28%), Positives = 146/275 (52%), Gaps = 11/275 (4%)

35  
 Query: 9 REFRLNRQFSLKQVASNELSVS QLSRFRGESDLSLTKFLGALEAIDL SISEFMDRVN KY 68  
 R R +Q S+ +A LS SQ+SRFRGES+++ ++ L L+ ++++I EF+ +K  
 Sbjct: 15 RRLRKGKQVSISFLADEYLSKSQISRFRGESEITCSRLNLLDKLNTIDEFVSAH SKT 74

40  
 Query: 69 QKSDQISLMSQMAQYHYQRDVAGLEKMISVEEGKLKDDSSDIRCRLNIVLFRGMICECDS 128  
 + +L+SQ + + +++V L K++ + KD R + +LF DS  
 Sbjct: 75 H-THFFTLLSQARKCYAEKNVVKLT KLL---KDYAHKDYE--RTMIKAILF-----SIDS 123

45  
 Query: 129 SRKMSEEDLCFLSDYLFQKDSWEISDYILIGNLYRYYNTRHICQLVKEVINQKEYYRDIY 188  
 S S+E+L L+DYLF+ + W + IL+GN R+ N + L KE++ Y  
 Sbjct: 124 SIAPSQEELTRLTDYLFKVEQWGYEIIILGNCSRFMNYNTLFLLTKEMVASFAYSEQNK 183

50  
 Query: 189 TNRNVEATLLNVVETLIERRALEEATFFLEKVEALLNNERNAYHRIILLYEKGFLAYAK 248  
 TN+ +V +N + I+ E + + + K++ LL +E N Y + + LY G+ +  
 Sbjct: 184 TNKMLVTQLSINCLII SIDHSCFEHSRYLINKIDLLLRDELNFYEKTVFLYVHGYYKLKQ 243

55  
 Query: 249 GDSRGIQSMKQAIFCFQAIGSKHHVENFQEHFNRV 283  
 + G + M+QA+ F+ +G +++EH+ ++  
 Sbjct: 244 EEMSGEEDMRQALQIFKYLGEDSLYYSYKEHYRQI 278

55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 193

60 A DNA sequence (GBSx0199) was identified in *S.agalactiae* <SEQ ID 629> which encodes the amino acid sequence <SEQ ID 630>. This protein is predicted to be permease. Analysis of this protein sequence reveals the following:

-264-

Possible site: 15

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

5 INTEGRAL Likelihood = -8.07 Transmembrane 217 - 233 ( 215 - 238)  
 INTEGRAL Likelihood = -7.96 Transmembrane 163 - 179 ( 158 - 185)  
 INTEGRAL Likelihood = -7.75 Transmembrane 71 - 87 ( 69 - 91)  
 INTEGRAL Likelihood = -7.22 Transmembrane 369 - 385 ( 356 - 389)  
 INTEGRAL Likelihood = -5.15 Transmembrane 279 - 295 ( 275 - 299)  
 INTEGRAL Likelihood = -4.88 Transmembrane 252 - 268 ( 250 - 270)  
 10 INTEGRAL Likelihood = -4.78 Transmembrane 140 - 156 ( 139 - 157)  
 INTEGRAL Likelihood = -3.56 Transmembrane 343 - 359 ( 340 - 367)  
 INTEGRAL Likelihood = -3.13 Transmembrane 40 - 56 ( 39 - 56)  
 INTEGRAL Likelihood = -2.28 Transmembrane 94 - 110 ( 92 - 112)

----- Final Results -----

15 bacterial membrane --- Certainty=0.4227(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

20 >GP:AAD36408 GB:AE001788 permease, putative [Thermotoga maritima]  
 Identities = 97/396 (24%), Positives = 194/396 (48%), Gaps = 15/396 (3%)

Query: 1 MNINGIKLLSSRAVSKLGDVFYDYGNSWIASMGGGLGQKILGIYQIVELLVSIVLNPFFGG 60  
 MN N + S VS +G Y + W+ S G + + G++ I L +I+++PF G  
 25 Sbjct: 1 MNRNLLLFASSFSVSLIGTRIQVALAWWLYSKTGSSEYV-GLFMISFLPAIIVSPFAG 59

Query: 61 ALADRFQRRKILLITDAICAIM---CFLLSFIGDDKVMVYGLIVANAILAVSNAFSSPAY 117  
 + DR RR ++++ D + ++ FL+ + + + + L++ +++V ++F +PA  
 30 Sbjct: 60 TVVDRHSRRNMVMVDILRGVLFMYLFLMEYFSELTMAL--LLIVTVLVSVFDSFFNPVAV 117

Query: 118 KSYIPEIVDKADIITYNANLETIVQIISVSSPVLGFLIFNFGIRITLIVDAITFLISFL 177  
 S +P++V K +++ N+ + + + P LG L+ G+ ++++++FLIS +  
 35 Sbjct: 118 DSLLPDLVRKENLVRANSILYRLKLNLSKILGPALGSLLLKVVGLAGVILINLSLFLISGI 177

Query: 178 FLYAIKVERVQLSKQEKVAIKNILADIADGFTYIKKEKEIMFFLIIAALLNTFLAMFNYL 237  
 F IKVE L K K +N+ DI YI+ + I+ +++ A++N F + L  
 40 Sbjct: 178 FEMFIKVEEKHLKKVSKE--RNMWQDIKSALLYIRSVRFILVTILVIAIMNFFTGSMHVL 235

Query: 238 LP-FTNSLLKTS GAYATILSISAIGSIIGALIARKI--KSSINSMLSMLVFSLSGLVIVMG 294  
 LP + L K+ Y T++S+ + G +I + I ++S+ ++ LV L V V  
 45 Sbjct: 236 LPEHVS KLKSEWVYGTLSMLSFGGLIVTFLMATIRTRASVKTGLNLVGYGLAVFVFA 295

Query: 295 FPSLFELPIWIPYSGSFLFNSLLTMFNIHFFSQVQIRVDEAYMGRVMSTIFTIAIMFMPI 354  
 W+ ++ FL T+FNI+ + +Q+ + E G++ S I ++ +P+  
 50 Sbjct: 296 MTGNH---WLMFAMYFLIGIFQTLFNINVTLLQLAIPEEMRGKIFSLISAVSFSLLPV 351

Query: 355 GTLFMTIFSFALSNVSFIVIGCAIAILGGLGFSYSK 390  
 F S ++ + I GG+ S +  
 55 Sbjct: 352 SYGFFGFLSSYVATAHIFITTSMALIAGGVLSLQR 387

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 631> which encodes the amino acid sequence <SEQ ID 632>. Analysis of this protein sequence reveals the following:

Possible site: 45

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

55 INTEGRAL Likelihood = -8.17 Transmembrane 172 - 188 ( 161 - 194)  
 INTEGRAL Likelihood = -8.07 Transmembrane 220 - 236 ( 218 - 242)  
 INTEGRAL Likelihood = -7.22 Transmembrane 311 - 327 ( 303 - 329)  
 INTEGRAL Likelihood = -5.26 Transmembrane 98 - 114 ( 96 - 118)  
 INTEGRAL Likelihood = -4.99 Transmembrane 347 - 363 ( 342 - 370)  
 60 INTEGRAL Likelihood = -4.62 Transmembrane 154 - 170 ( 151 - 171)  
 INTEGRAL Likelihood = -4.25 Transmembrane 284 - 300 ( 281 - 306)  
 INTEGRAL Likelihood = -3.66 Transmembrane 378 - 394 ( 378 - 396)  
 INTEGRAL Likelihood = -3.56 Transmembrane 74 - 90 ( 73 - 92)  
 65 INTEGRAL Likelihood = -2.39 Transmembrane 50 - 66 ( 49 - 66)

## ----- Final Results -----

bacterial membrane --- Certainty=0.4270(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the databases:

>GP:AAD36408 GB:AE001788 permease, putative [Thermotoga maritima]

Identities = 85/345 (24%), Positives = 171/345 (48%), Gaps = 8/345 (2%)

10

Query: 40 SLSLVAVYQSLQSVIGVLFNLFQGGVIADSFKRKKIITTNILCGTACLVLSFLTKEQWLV 99

S V ++ + ++ + F G + D R+ +++ +IL G + L + L

Sbjct: 36 SSEYVGLFMISFLPAIIVSPFAGTVDRHSRRNMVMVDILRGVLFMYLFIMEYFSELT 95

15

Query: 100 YAIVL-TNVILAFMSAFSSPSYKAFTKEIVKDSISQLNSLLETTSTVIKVTVPMAIFL 158

A++L V+++ +F +P+ + ++V+K+++ + NSL + K+ P + L

Sbjct: 96 MALLLIVTVLVSVFDSFFNPAPVDSLPLDVRKENLVRANSYRLKLNLSKILGPALGSLL 155

20

Query: 159 YKLLGIHGVLLLDGLSFLIAALLISFILPVNDEVVIKEKVTIREIFNDLKIGFKYVYSHK 218

K++G+ GV+L++ LSFLI+ + FI +E +K+ R ++ D+K Y+ S +

Sbjct: 156 LKVVGLAGVILINLSLFLISGIFEMFIKV--EEKHLKKVSKERNMQDIKSALLYIRSVR 213

25

Query: 219 SIFIITVLSALVNFFFLAAYNLLLPYSNQMFGEISTGLYGTFLTAEAGGFIGAILSGFVN 278

I + ++ A++NFF + ++LLP G+ S +YGT ++ + GG I L +

Sbjct: 214 FILVTILVIAIMNFFTGSMDHLLPEHVSGLGK-SEWVYGTLMSSLSFGGLIVTFLMATIR 272

Query: 279 KELSSMRLILFLSLSGMLMLLAPPFYIMFHNAIILALSPALFSLFLSIFNIQFFSLVQKD 338

S L L L GL + + + M N ++ L +F ++FNI +L+Q

Sbjct: 273 TRASVKTGLNLVGYGLAVFV----FAMTGNHWMFMAMYFLIGIFQTLFNINVTLLQLA 328

30

Query: 339 VDNDFLGRVFGIIFTITILFMPIGTGFFSVALNPNNNSFNLFIIIGS 383

+ + G++F +I ++ +P+ GFF + + ++FI S

Sbjct: 329 IPEEMRGKIFSLISAVSFSLPVSYGFFGFLSSYVATAHIFITTS 373

An alignment of the GAS and GBS proteins is shown below:

35

Identities = 136/379 (35%), Positives = 229/379 (59%), Gaps = 6/379 (1%)

Query: 8 LLSSRAVSKLGDVIFYDYNSTWIASMGLGQKILGIYQIVELLVSIVLNPFGGALADRFQ 67

L+ S+ + ++GDV +D+ N+T++A + ++ +YQ +E ++ ++ N FGG +AD F+

Sbjct: 11 LVYSKVIYRIGDVMFDFANNTFLAGLNPAASLSLVAVYQSLQSVIGVLFNLFQGGVIADSF 70

40

Query: 68 RRKILLITDAICAIMCFLLSFQGGDKVMVYGLIVANAILAVSNAFSSPAYKSYIPEIVDK 127

R+KI++ T+ +C C +LSF+ ++ +VY +++ N ILA +AFSSP+YK++ EIV K

Sbjct: 71 RKKIIITTNILCGTACLVLSFLTKEQWLVYAIVLTNVILAFMSAFSSPSYKAFTKEIVKK 130

45

Query: 128 ADIITYNANLETIVQIISVSSPVLGFLIFNFGIRITLIVDAITFLISFLFLYAIKVERV 187

I N+ LET +I V+ P++ ++ GI L++D ++FLI+ L + I

Sbjct: 131 DSISQLNSLLETTSTVIKVTVPMAIFLYKLLGIHGVLLLDGLSFLIAALLISFILPVND 190

50

Query: 188 QLSKQEKVAIKNILADIADGFTYIKKEKEIMFFLIIAALLNTFLAMFNLLPFTNSLLK- 246

++ +EKV I+ I D+ GF Y+ K I +++AL+N FLA +N LLP++N +

Sbjct: 191 EVVIKEKVTIREIFNDLKIGFKYVYSHKSIFIITVLSALVNFFFLAAYNLLLPYSNQMFGE 250

Query: 247 -TSGAYATILSISAIGSIIGALIARKIKSSINSMLSVFSSLGVIVMGFPS---LFELP 302

++G Y T L+ AIG IGA+++ + ++SM +L S G+++M P +F

55

Sbjct: 251 ISTGLYGTFLTAEAGGFIGAILSGFVNKELSSMRLILFLSLSGMLMLLAPPFYIMFHNA 310

Query: 303 IWIPYSGSFLFNLLTMFNIHFFSQVQIRVDEAYMGRVMSTIFTIIMFMPIGTFLMTIF 362

I + S + LF+ L++FNI FFS VQ VD ++GRV IFTI I+FMPIGT F ++

Sbjct: 311 IILALSPA-LFSLFLSIFNIQFFSLVQKDVDNDFLGRVFGIIFTITILFMPIGTGFFSVA 369

60

Query: 363 SFALSNVSFIVIGCAIAIL 381

++ + +IG I L

Sbjct: 370 LNPNNNSFNLFIIIGSCITTL 388

-266-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 194

A DNA sequence (GBSx0200) was identified in *S. agalactiae* <SEQ ID 633> which encodes the amino acid sequence <SEQ ID 634>. This protein is predicted to be membrane permease OpuCD. Analysis of this protein sequence reveals the following:

```

Possible site: 46
>>> Seems to have an uncleavable N-term signal seq
  INTEGRAL    Likelihood = -5.68    Transmembrane    91 - 107 ( 88 - 110)
  INTEGRAL    Likelihood = -4.30    Transmembrane    15 - 31 ( 9 - 37)
  INTEGRAL    Likelihood = -3.72    Transmembrane    72 - 88 ( 72 - 88)
  INTEGRAL    Likelihood = -3.19    Transmembrane   124 - 140 ( 123 - 142)

----- Final Results -----
bacterial membrane --- Certainty=0.3272(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8509> which encodes amino acid sequence <SEQ ID 8510> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 1
McG: Discrim Score:    -10.69
GvH: Signal Score (-7.5): -3.79
Possible site: 39
>>> Seems to have no N-terminal signal sequence
ALOM program    count: 5 value: -9.02 threshold: 0.0
  INTEGRAL    Likelihood = -9.02    Transmembrane    35 - 51 ( 25 - 53)
  INTEGRAL    Likelihood = -5.68    Transmembrane   151 - 167 ( 148 - 170)
  INTEGRAL    Likelihood = -4.30    Transmembrane    75 - 91 ( 69 - 97)
  INTEGRAL    Likelihood = -3.72    Transmembrane   132 - 148 ( 132 - 148)
  INTEGRAL    Likelihood = -3.19    Transmembrane   184 - 200 ( 183 - 202)
  PERIPHERAL  Likelihood = 2.17      58
modified ALOM score: 2.30

*** Reasoning Step: 3

----- Final Results -----
bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAF91342 GB:AF249729 membrane permease OpuCD [Listeria monocytogenes]
Identities = 104/154 (67%), Positives = 133/154 (85%)

Query: 3    IANVIQTIPSLAMISIIIMGLGLGIKTVVATVFLYSLLPITNTYTGTIRNVSDLLDAAK 62
           IAN+IQTIP+LAM++++ML +GLG TVV ++FLYSLLPi+ NTYTGTIRNV LL++ K
Sbjct: 60   IANIIQTIPALAMLAVALMLIMGLGTNTFVLSLFLYSLLPILKNTYTGTIRNVGALLESGK 119

Query: 63   GMGMTKRQRLFMVELPLSISVIMAGLRNALVVAIGITAIGAFVGGGGLGDIIRGTNATN 122
           MGMTK Q L ++E+PL++SVIMAG+RNALV+AIG+ AIG FVG GGLGDII+RGTNATN
Sbjct: 120  AMGMTKWQVLRLEIMPLALSVMAGIRNALVIAIGVAAIGTFVGAGGLGDIIVRGTNATN 179

Query: 123  GGAIILAGSLPTALMAIFSDLILGGIQRMLEPRK 156
           G AIILAG++PTA+MAI +D++LG ++R L P K
Sbjct: 180  GTAILAGAIPTAVMAILADVLLGWVERTLNPKV 213

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 635> which encodes the amino acid sequence <SEQ ID 636>. Analysis of this protein sequence reveals the following:

-267-

Possible site: 49

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -9.24 Transmembrane 39 - 55 ( 31 - 59)  
 INTEGRAL Likelihood = -7.17 Transmembrane 190 - 206 ( 188 - 211)  
 INTEGRAL Likelihood = -4.62 Transmembrane 93 - 109 ( 75 - 110)  
 INTEGRAL Likelihood = -3.66 Transmembrane 76 - 92 ( 75 - 92)  
 INTEGRAL Likelihood = -2.87 Transmembrane 221 - 237 ( 220 - 237)  
 INTEGRAL Likelihood = -2.44 Transmembrane 168 - 184 ( 165 - 184)

----- Final Results -----

bacterial membrane --- Certainty=0.4694 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAD45530 GB:AF162656 choline transporter [Streptococcus pneumoniae]  
 Identities = 344/508 (67%), Positives = 425/508 (82%), Gaps = 2/508 (0%)

Query: 13 MPSLFVTFQNRFNELAAALGEHLQISLLSLMIALIGVPLAALLSRKRWSDIMLQVTGV 72  
 M +L TFQ+RF++WL AL +HLQ+SLL+L++A+L+ +PLA L ++ +D +LQ+ G+  
 Sbjct: 1 MTNLIATFQDRFSDWLTAQSQHLQSLTLLLAILLAIPAVFLRYHEKLADWVLQIAGI 60

Query: 73 FQTIPSLALLGLFIPLMGIGTLPALTALVIYAIFFPILQNTITGLNGIDPSLVEAGIAFGM 132  
 FQTIPSLALLGLFIPLMGIGTLPALTALVIYAIFFPILQNTITGL GIDP+L EAGIAFGM  
 Sbjct: 61 FQTIPSLALLGLFIPLMGIGTLPALTALVIYAIFFPILQNTITGLKGIDPNLQEAGIAFGM 120

Query: 133 TKWERLKTFEIPIAMPVIMSGVRTSAVMIIGTATLASLIGAGGLGSFILLGIDRNNANLI 192  
 T+WERLK FEIP+AMPVIMSG+RT+AV+IIGTATLA+LIGAGGLGSFILLGIDRNN+LI  
 Sbjct: 121 TRWERLKKFEIPLAMPVIMSGIRTAAVLIIGTATLAALIGAGGLGSFILLGIDRNNASLI 180

Query: 193 LIGAISSALLAIIFNSLLQYLEKASLRIMISFGITLLALLASYTPMALSQFSKGDITVV 252  
 LIGA+SSA+LAI FN LL+ +EKA LR I F + L L SY+P L Q K K+ +V  
 Sbjct: 181 LIGALSSAVLAIAFNFLKVMKAKLRTIFSGFALVALLGLSYSPALLVQ--KEKENLV 238

Query: 253 IAGKLGAEPDILINLYKELIEDQSDISVELKSNFGKTSFLYEALKSGDIDMYEFTGTIT 312  
 IAGK+G EP+IL N+YK LIE+ + ++ +K NFGKTSFLYEALK GDID+YFEFTGT+T  
 Sbjct: 239 IAGKIGPEPEILANMYKLLIEENTSMTATVKPNFGKTSFLYEALKKGDDIDIYFEFTGTVT 298

Query: 313 SLLLRDKPPLSNDPKQVYEDAKKGIKQDKLTLLKPFAYQNTYAVAMPEKLAKEYQIETI 372  
 SLL+ P +S++P+QVY+ A+ GIAKQD L LKP +YQNTYAVA+P+K+A+EY ++TI  
 Sbjct: 299 ESLLQPSPKVSHEPEQVYQVARDGIKQDHLAYLKPMYQNTYAVAVPKKIAQEYGLKTI 358

Query: 373 SDLKAHADTLKAGFTLEFKDRADGYKGMQSQYGLQLSVATMEPALRYQAIQSGDIQVTD 432  
 SDLK LKAGFTLEF DR DG KG+QS YGL L+VAT+EPALRYQAIQSGDIQ+TDA  
 Sbjct: 359 SDLKKVEGQLKAGFTLEFNREDGNKGLQSMYGLNLNVATIEPALRYQAIQSGDIQITDA 418

Query: 433 YSTDAEITKYHLKVLKDDKQLFPPYQGAPLMKTSLLTKHPELKGILNQLAGKITEKEMQD 492  
 YSTDAE+ +Y L+VL+DDKQLFPPYQGAPLMK +LL KHEP+ +LN LAGKITE +M  
 Sbjct: 419 YSTDABELERYDLQVLEDDKQLFPPYQGAPLMKEALLKKHPELRLVNTLAGKITESQMSQ 478

Query: 493 MNYEVSVKGADANKVARDYLLKTGLIQK 520  
 +NY+V V+G A +VA+++L + GL++K  
 Sbjct: 479 LNYQVGVEGKSAKQVAKQFLQEQGLLKK 506

An alignment of the GAS and GBS proteins is shown below:

Identities = 53/148 (35%), Positives = 93/148 (62%), Gaps = 1/148 (0%)

Query: 3 IANVIQTIPSLAMISIIIMGLGLGIKTIVVATVFLYSLLPITNTYTGIRNVSDLLDAK 62  
 + V QTIPSLA++ + + +G+G V + +Y++ PI+ NT TG+ +D L++A  
 Sbjct: 69 VTGVFQTIPSLALLGLFIPLMGIGTLPALTALVIYAIFFPILQNTITGLNGIDPSLVEAGI 128

Query: 63 GMGMTKRQRLFMVELPLSISVIMAGLRNALVVAIGITAIGAFVGGGGLDIIIRGTNATN 122  
 GMTK +RL E+P+++ VIM+G+R + V+ IG + + +G GGLG I+ G + N  
 Sbjct: 129 AFGMTKWERLKTFEIPIAMPVIMSGVRTSAVMIIGTATLASLIGAGGLGSFILLGIDRNN 188

Query: 123 GGAIILAGSLPTALMAIFSDLILGGIQR 150

-268-

+IL G++ +AL+AI + +L +++  
 Sbjct: 189 AN-LILIGAISALLAIIFNSLLQYLEK 215

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 195

A DNA sequence (GBSx0201) was identified in *S.agalactiae* <SEQ ID 637> which encodes the amino acid sequence <SEQ ID 638>. This protein is predicted to be choline transporter-related. Analysis of this protein sequence reveals the following:

Possible site: 44  
 >>> May be a lipoprotein  
 INTEGRAL Likelihood = -3.03 Transmembrane 306 - 322 ( 306 - 327)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2211(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9517> which encodes amino acid sequence <SEQ ID 9518> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB15386 GB:Z99121 glycine betaine/carnitine/choline ABC  
 transporter (osmoprotectant-binding protein) [Bacillus subtilis]  
 Identities = 168/303 (55%), Positives = 224/303 (73%), Gaps = 1/303 (0%)  
 Query: 2 LKKS HFLQIFTLCLALLTISGCQLTDTKKSGHTTIKVAQSSSTESSIMANIITELIHHEL 61  
 + K +L F L +L + GC L + TIK+ AQS TES I+AN+I +LI H+  
 Sbjct: 1 MTKIKWLGAFAFVFMV-LGGCSLPGLGGASDDTIKIGAQSMTSEIVANMIAQLIEHDT 59  
 Query: 62 GYNTTLISNLGSSSTVTHQALLRGDADIAATRYTGTDTITGLGLKAVKDPKEASKIVKTEF 121  
 NT L+ NLGS+ V HQA+L GD DI+ATRY+GTD+T TLG +A KDPK+A IV+ EF  
 Sbjct: 60 DLNTALVKNLGSNVVQHQAMLGGDIDISATRYSGTDLTSTLGKEAEKDPKALNIVQNEF 119  
 Query: 122 QKRYNQTYPTYGFSPTYAFMVTKEFARQNKITKISDLKKLSTTMKAGVDSSWMNREGDG 181  
 QKR++ W+ +YGF +TYAF VTK+FA + I +SDLKK ++ K GVD++W+ R+GDG  
 Sbjct: 120 QKRFSYKWFDSYGFNTYAFVTYKFAEKEHINTVSDLKKNASQYKLGVDNAWLKRKG DG 179  
 Query: 182 YTDFAKTYGFEFHSIYPMQIGLVYDAVESNKMQSVLGYSTDGRISYDLEILRDDKKFFP 241  
 Y F TYGFEF YPMQIGLVYDAV++ KM +VL YSTDGRI +YDL+IL+DDK+FFP  
 Sbjct: 180 YKGFVSTYGFEGFTYPMQIGLVYDAVKNKGMDAVLAYSTDGRIKAYDLKILKDDKRFFP 239  
 Query: 242 PYEASMVVNNSIIKKDPKLLKLLHRLDGKINLKTMONLNYMVDKLEPSVVAKEFLEKN 301  
 PY+ S V+ ++K+ P+L+ ++++L G+I+ +TMQ LNY VD KL EPSVVAKEFLEK+  
 Sbjct: 240 PYDCSPVIEKVLKEHPELEGVINKLIGQIDTETMQELNVEVDGKLKEPSVVAKEFLEKH 299  
 Query: 302 HYF 304  
 HYF  
 Sbjct: 300 HYF 302

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8511> and protein <SEQ ID 8512> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: 22 Crend: 5  
 McG: Discrim Score: 10.26  
 GvH: Signal Score (-7.5): -4.19

```

Possible site: 44
>>> May be a lipoprotein
ALOM program count: 0 value: 8.65 threshold: 0.0
PERIPHERAL Likelihood = 8.65 66
modified ALOM score: -2.23

```

```

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

```

EGAD|109208| glycine betaine/carnitine/choline ABC Insert characterized
SP|O32243|OPCC_BACSU    GLYCINE    BETAINE/CARNITINE/CHOLINE-BINDING    PROTEIN    PRECURSOR
(OSMOPROTECTANT-BINDING
PROTEIN). Insert characterized
GP|2635894|emb|CAB15386.1||Z99121 glycine betaine/carnitine/choline ABC transporter
(osmoprotectant-binding protein) Insert characterized
PIR|E69670|E69670 glycine betaine/carnitine/choline ABC transporter (osmoprotec) opuCC -
Insert characterized

```

```

EGAD|109208|BS3376(15 - 302 of 303) glycine betaine/carnitine/choline ABC {Bacillus subtilis} SP|O32243|OPCC_BACSU GLYCINE BETAINE/CARNITINE/CHOLINE-BINDING PROTEIN PRECURSOR (OSMOPROTECTANT-BINDING PROTEIN). GP|2635894|emb|CAB15386.1||Z99121 glycine betaine/carnitine/choline ABC transporter (osmoprotectant-binding protein) {Bacillus subtilis} PIR|E69670|E69670 glycine betaine/carnitine/choline ABC transporter (osmoprotec) opuCC - Bacillus subtilis
%Match = 33.5
%Identity = 56.2 %Similarity = 75.3
Matches = 162 Mismatches = 71 Conservative Sub.s = 55

```

402            432            462            492            522            552            582            612  
 LTDTKKS~~G~~H~~T~~~~T~~I~~K~~VAAQSSTESSIMANIITELIHHELGYNTLTISNLGSSTVTHQALLRGDADIAATRYTGTGITGLGL  
 | : |||: ||| ||| |:||: |::| : || || |::| |||: ||| ||| :|||: |||:  
 LPGLGGASDDTIKIGAQSMTESEIVANMIAQLIEHDTDLNLTALVKNLGSNVYQHQAAMLGDDIDISATRYSGTDLTSLGK  
                40            50            60            70            80            90            100

642            672            702            732            762            792            822            852  
KAVKDPKEASKIVKTEXQKRYNQTWYPTYGFSDTYA FMTVTKEFARQNKITIKISDLKKLSTTMKAGVDSSWMNREGDGYTD  
:| ||||:| |:: |: ||| :||| ||::| :| :||| :: | ||::| :|||  
EAEKDKPKALNI VONEFOKRFSYKWFD SYGF DNTYAFTVTKKFAEK EHI NTVSDLKKNASQYKLGVDNAWLKRKG DG YGK  
120            130            140            150            160            170            180

882           912           942           972           1002           1032           1062           1092  
FAKTYGFEFSHIYPMQIGLVYDAVESNKMQSVLGYSTDGRISSYDLEILRDDKKFFPPYEASMVVNNSIIKKDKPLKKLL  
| | | | | | | | | | | | | | : | : | | | | | : | : | : | : | : | : | : | : | : | :  
FVSTYGFEFGTTYPMQIGLVYDAVNGKMDAVLAYSTDGRIKAYDLKILKDDKRFFPPYDCSPVIPEKVLKEHPELEGVI

200           210           220           230           240           250           260

1122      1152      1182      1212      1242      1272      1302      1332  
HRLDGKINLKTMQNLNMYMDDKLLEPSVVAQKFLEKNHYFRGDK\*MKQMNTFQQFIYYFQHNGSYILEQFIHHFLISVYG  
:: | :| : | || | | | | | | :|| :||  
NKLIQGIDTETMQELNYEVDGKLKEPSVVAKEFLEKHHYFD

                280         290         300

SEQ ID 8512 (GBS23) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 8; MW 35kDa).

-270-

The GBS23-His fusion product was purified (Figure 194, lane 9) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 251). These tests confirm that the protein is immunoaccessible on GBS bacteria.

### Example 196

A DNA sequence (GBSx0202) was identified in *S.agalactiae* <SEQ ID 639> which encodes the amino acid sequence <SEQ ID 640>. This protein is predicted to be membrane permease OpuCB (opuBB). Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.66	Transmembrane	25 - 41 ( 18 - 45)
INTEGRAL	Likelihood = -7.96	Transmembrane	182 - 198 ( 174 - 202)
INTEGRAL	Likelihood = -4.83	Transmembrane	61 - 77 ( 57 - 95)
INTEGRAL	Likelihood = -4.09	Transmembrane	78 - 94 ( 78 - 95)
INTEGRAL	Likelihood = -1.22	Transmembrane	134 - 150 ( 134 - 150)

----- Final Results -----

bacterial membrane	---	Certainty=0.4864 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF91340 GB:AF249729 membrane permease OpuCB [Listeria monocytogenes]

Identities = 121/208 (58%), Positives = 160/208 (76%)

Query: 1	MVNFLSQYGMQILVKTWQVYISFFAIALGIAIAVPLGVVLTRFPKVAKIIIAIASMLQT	60
	+V F + G +LV+TW+ ++IS A+ LGIA+AVP G++LTR PKVA +I + S+LQT	
Sbjct: 4	IVTFFQENGHNLLVQTWQHLFISLSAVILGIAVAVPTGILLTRSPKVANFVIGVSVLQT	63
Query: 61	IPSLALLALMIPFLFGIGKIPAIVALFIYSLLPILRNTYIGMNNVNPTLKDKCAKGMGMKPI	120
	+PSLA+LA +IP G+G +PAI+ALFIY+LLPILRNT+IG+ V+ L + +GMGM	
Sbjct: 64	VPSLAIALAFIIPFLGVGTLPAILALFIYALLPILRNTFIGVRGVDKNLIESGRGMGMTNW	123
Query: 121	QSIFQVELPLATPIIMAGIRLSITIYVIAWATLASYGAGGLGDLIFSGNLNFQSKLILGG	180
	Q I VE+P + +IMAGIRLS +YVIAWATLASYGAGGLGD IF+GLNL++ LILGG	
Sbjct: 124	QLIVNVEIPNSISVIMAGIRLSAVYVIAWATLASYGAGGLGDFIFNGLNLYRPDLILGG	183
Query: 181	TIPVIILSLIIDYLLGLELTALTPRTTR	208
	IPV IL+L++++ LG LE LTP+ R	
Sbjct: 184	AIPVTILALVVEFALGKLEYRLTPKAIR	211

A related GBS gene <SEQ ID 8513> and protein <SEQ ID 8514> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 0

McG: Discrim Score: -9.08

GvH: Signal Score (-7.5): -1.86

Possible site: 37

>>> Seems to have no N-terminal signal sequence

ALOM program count: 5 value: -8.60 threshold: 0.0

INTEGRAL	Likelihood = -8.60	Transmembrane	25 - 41 ( 18 - 45)
INTEGRAL	Likelihood = -7.96	Transmembrane	182 - 198 ( 174 - 202)
INTEGRAL	Likelihood = -4.83	Transmembrane	61 - 77 ( 57 - 95)
INTEGRAL	Likelihood = -4.09	Transmembrane	78 - 94 ( 78 - 95)
INTEGRAL	Likelihood = -1.22	Transmembrane	134 - 150 ( 134 - 150)
PERIPHERAL	Likelihood = 2.70		156
modified ALOM score: 2.22			

\*\*\* Reasoning Step: 3





-272-

Query: 125 LVELPEEYLD RYPSELSSGGQQQRIGVIRALAADQDIILMDEPFGALDPITREGIQDLVKS 184  
 LV+LPEE+LD RYP ELSSGGQQQRIGV+RALAA+Q++ILMDEPFGALDPITR+ +Q+ K+  
 Sbjct: 121 LVDLPEEFLDRYPYELSSGGQQQRIGVLRALAAEQNLILMDEPFGALDPITRDSLQEEFKN 180

5 Query: 185 LQEE MGKTIILVTHDMDEALKLATKII VMDNGKMVQEGTPNDLLHHPATSFVEQMIGEER 244  
 LQ+E+GKTII VTHDMDEA+KLA +I++M +G++VQ TP+++L +PA SFVE IG++R  
 Sbjct: 181 LQKELGKTIIFVTHDMDEAIKLAD RIVIMKDGEIVQFDTPEILRN PANSFVEDFIGKDR 240

10 Query: 245 LLHAQADITPVKQIMLNPNVSITA EKTILTEAITLMRQKRVD SLLVTDNGKLI-GFIDLES 303  
 L+ A+ D+T V QIM NPVSITA+K+L AIT+M++KRVD+LLV D G ++ GFID+E  
 Sbjct: 241 LIEAKPDVTQVAQIMNTPNVSITADKSLQAAITVMKEKRVDTLLVVD EGNVLKGFIDVEQ 300

15 Query: 304 LSSKYKKDR LVS DILKHTDFYVMEDDLRNTAERILKGLKYAPVVDH ENNLKGIVTRAS 363  
 + + V DI++ FYV ED LLR+T +RILK G KY PVVD + L GIVTRAS  
 Sbjct: 301 IDLNRR TATSVMDIIEKNV FVYEDTLLRDTVQRILKRGYKYIPVVDKDKRLVGIVTRAS 360

20 Query: 364 LVDMLYDIIWGDTE--TEDQ 381  
 LVD++YD IWG E TE+Q  
 Sbjct: 361 LVDIVYDSIWGTLEDATENQ 380

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 643> which encodes the amino acid sequence <SEQ ID 644>. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3619(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 102/237 (43%), Positives = 165/237 (69%), Gaps = 1/237 (0%)

35 Query: 6 IIEYQNINKVYGENVAVEDINLKIYPGDFVCFIGTSGSGKTTLMRMVNHMLKPTNGTLLF 65  
 +I + N++K +G+ +++ +I +F +G SGSGKTTL++M+N +++P++G +L  
 Sbjct: 1 MIRFNNVSKTFGQTKVLQEQTFQINDREFFVLVGPSSGSGKTTLLKMINCLIEPSSGDILL 60

40 Query: 66 KGKDISTINPIELRRRIGYVIONIGLMPHMTIYENIVLPKLLKWSEEAKRAKARELIK 125  
 + ++ E+R IGYV+Q I L P++T+ ENI ++P++ +WS E R K EL+  
 Sbjct: 61 NNVPTQTELDLREMLRSIGYVLQQIALFPNLTVAENIAIIPEMKQWSAEEIRQKTEELLDK 120

45 Query: 126 VELP-EEYLD RYPSELSSGGQQQRIGVIRALAADQDIILMDEPFGALDPITREGIQDLVKS 184  
 V LP ++YLD RYPS+LSGG+QQRIG++RA+ + I+LMDEPF ALDPI+R+ +Q+L+ S  
 Sbjct: 121 VGLPAKYLD RYPSDLSGGEQQRIGIVRAIISHPKILMDEPF S ALDPISRKQLQELMLS 180

50 Query: 185 LQEE MGKTIILVTHDMDEALKLATKII VMDNGKMVQEGTPNDLLHHPATSFVEQMIG 241  
 L +E TI+ VTHD+DEA+KL ++ +++ G++VQ P + HPA +FV + G  
 Sbjct: 181 LHKEFDMTIVFVTHDIDEAIKLGD R VAILNEGEIVQLDRPEMIKTHPANAFVNNLFG 237

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 198

A repeated DNA sequence (GBSx0212) was identified in *S.agalactiae* <SEQ ID 645> which encodes the amino acid sequence <SEQ ID 646>. Analysis of this protein sequence reveals the following:

55 Possible site: 24  
 >>> Seems to have no N-terminal signal sequence

60 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4736(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-273-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 199

A DNA sequence (GBSx0213) was identified in *S.agalactiae* <SEQ ID 647> which encodes the amino acid sequence <SEQ ID 648>. Analysis of this protein sequence reveals the following:

10 Possible site: 38  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.06 Transmembrane 18 - 34 ( 18 - 34)  
 ----- Final Results -----  
 15 bacterial membrane --- Certainty=0.1426(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 A related GBS gene <SEQ ID 8515> and protein <SEQ ID 8516> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: 20 Crend: 5  
 Sequence Pattern: CQMN  
 SRCFLG: 0  
 25 McG: Length of UR: 19  
 Peak Value of UR: 2.60  
 Net Charge of CR: 3  
 McG: Discrim Score: 7.77  
 GvH: Signal Score (-7.5): -4.89  
 30 Possible site: 25  
 >>> May be a lipoprotein  
 Amino Acid Composition: calculated from 21  
 ALOM program count: 0 value: 13.21 threshold: 0.0  
 PERIPHERAL Likelihood = 13.21 115  
 35 modified ALOM score: -3.14  
 \*\*\* Reasoning Step: 3  
 ----- Final Results -----  
 40 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

45 ORF01527(346 - 465 of 1095)  
 EGAD|7398|7198(2 - 41 of 47) lysis protein for colicin e9 precursor {Escherichia coli}  
 EGAD|41475|43808 lysis protein { } SP|P13344|LYS5\_ECOLI LYSIS PROTEIN FOR COLICIN E5  
 PRECURSOR. GP|40543|emb|CAA33861.1||X15857 lysis protein (AA 1-47) {Enterobacteriaceae}  
 50 GP|144373|gb|AAA98053.1||M30445 colicin release protein {Plasmid ColE5-099}  
 PIR|JQ0330|JQ0330 colicin E5 lysis protein precursor - Escherichia coli plasmid ColE5-099  
 %Match = 3.7  
 %Identity = 35.0 %Similarity = 52.5  
 Matches = 14 Mismatches = 19 Conservative Sub.s = 7  
 55 135 165 195 225 255 285 315 345  
 YIYFFHCRRRIYIIININY\*FN\*GI\*NIQMIFCLHVKTPTKIKIRENFVILKLIL\*CW\*IIVNFIIYLIYKIYILRKENMMR

M

5                    375                    405                    435                    465                    495                    525                    555                    585  
 KYIKWLIPISIFGMILGGCQMNSEHKIQSNEVKNSKQSEVKKDKKMTKKEQLAYLKEHEQEIIDYVKLHNNQIESVQFDW  
 | | | : : : | | | | : | | | : | :  
 KKITWIILLLLAAILAACQANYIHDVQGGTVSPSSSAELTGLATQ  
    20                    30                    40

10    SEQ ID 8516 (GBS389) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 74 (lane 6; MW 18kDa).

The GBS389-His fusion product was purified (Figure 214, lane 4) and used to immunise mice. The resulting antiserum was used for FACS (Figure 313), which confirmed that the protein is immunoaccessible on GBS bacteria.

15    Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 200

A DNA sequence (GBSx0214) was identified in *S.agalactiae* <SEQ ID 649> which encodes the amino acid sequence <SEQ ID 650>. Analysis of this protein sequence reveals the following:

20       Possible site: 19  
        >>> Seems to have no N-terminal signal sequence  
  
        ----- Final Results -----  
                  bacterial cytoplasm --- Certainty=0.3766(Affirmative) < succ>  
 25                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

30    Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 201

A DNA sequence (GBSx0215) was identified in *S.agalactiae* <SEQ ID 651> which encodes the amino acid sequence <SEQ ID 652>. Analysis of this protein sequence reveals the following:

35       Possible site: 46  
        >>> Seems to have no N-terminal signal sequence  
  
        ----- Final Results -----  
                  bacterial cytoplasm --- Certainty=0.3882(Affirmative) < succ>  
 40                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45    Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 202**

A DNA sequence (GBSx0216) was identified in *S.agalactiae* <SEQ ID 653> which encodes the amino acid sequence <SEQ ID 654>. This protein is predicted to be lectin, alpha subunit precursor. Analysis of this protein sequence reveals the following:

```

5   Possible site: 47
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.0653(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 203**

A DNA sequence (GBSx0217) was identified in *S.agalactiae* <SEQ ID 655> which encodes the amino acid sequence <SEQ ID 656>. Analysis of this protein sequence reveals the following:

```

20   Possible site: 41
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
25      bacterial cytoplasm --- Certainty=0.6569(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 204**

A DNA sequence (GBSx0218) was identified in *S.agalactiae* <SEQ ID 657> which encodes the amino acid sequence <SEQ ID 658>. Analysis of this protein sequence reveals the following:

```

35   Possible site: 27
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
40      bacterial cytoplasm --- Certainty=0.5736(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 205**

A DNA sequence (GBSx0219) was identified in *S.agalactiae* <SEQ ID 659> which encodes the amino acid sequence <SEQ ID 660>. Analysis of this protein sequence reveals the following:

```

Possible site: 52
5  >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -13.11    Transmembrane  146 - 162 ( 138 - 170)
    INTEGRAL    Likelihood = -12.90    Transmembrane   13 - 29 (   9 - 32)
    INTEGRAL    Likelihood = -9.50     Transmembrane  108 - 124 ( 104 - 129)
    INTEGRAL    Likelihood = -7.75     Transmembrane   40 - 56 (  33 - 61)
10  INTEGRAL    Likelihood = -6.64     Transmembrane  177 - 193 ( 170 - 195)
    INTEGRAL    Likelihood = -3.35     Transmembrane   77 - 93 (  77 - 97)

----- Final Results -----
15      bacterial membrane --- Certainty=0.6243(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8517> which encodes amino acid sequence <SEQ ID 8518> was also identified.

20 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 206**

25 A DNA sequence (GBSx0220) was identified in *S.agalactiae* <SEQ ID 661> which encodes the amino acid sequence <SEQ ID 662>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have no N-terminal signal sequence

30  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2374(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

35 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAB89623 GB:AE000990 repressor protein [Archaeoglobus
    fulgidus]
    Identities = 34/62 (54%), Positives = 46/62 (73%)

40  Query: 11 LKQVREDIGMTQQELAIRIGVRRETIGHLENNRYNPSLEMALKIVKIFDMKIEDIFQLRK 70
      +K+ R   MTQ+ELA R+GVRRETI LE  +YNPSL++A KI ++F+ KIEDIF  +
    Sbjct: 5  IKEFRAKFNMTQEELAKRVGVRRETIVFLEKGYNPSLKLAYKIARVFNKIEDIFIFDE 64

    Query: 71 ED 72
45      E+
    Sbjct: 65 EE 66

```

There is also homology to SEQ ID 412.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 207**

A DNA sequence (GBSx0221) was identified in *S.agalactiae* <SEQ ID 663> which encodes the amino acid sequence <SEQ ID 664>. Analysis of this protein sequence reveals the following:

Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3794 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB61817 GB:AL133236 putative acetyl transferase [Streptomyces  
coelicolor A3(2)]

Identities = 30/97 (30%), Positives = 52/97 (52%), Gaps = 1/97 (1%)

Query: 82 VGMLNIVTLARADMQWGE L GYVFHNQFWSNGYAFESILALLNSTYEKLG FHHIEAQITPG 141

VGM ++ + Q GE+ Y+ H + W G E +LL+ +++ G H I A P

Sbjct: 72 VGMGDLHVRSHTRQ-GEISYIVHPRVWGQIGTEIGRSLLSLGFDRWGLHRIRATCDPR 130

Query: 142 NERSEKLVRR LGLTYETTRKDFS FENGKWTDKLIYSI 178

N+ S +++ +LG+TYE + ++ W D L++SI

Sbjct: 131 NQASSRVLTKLGMTYEGRRHRTAWIRDGWRDSLVSFI 167

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 208**

A DNA sequence (GBSx0222) was identified in *S.agalactiae* <SEQ ID 665> which encodes the amino acid sequence <SEQ ID 666>. This protein is predicted to be p20 protein. Analysis of this protein sequence reveals the following:

Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1044 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA30415 GB:X07542 P20 (AA 1-178) [Bacillus licheniformis]

Identities = 56/175 (32%), Positives = 94/175 (53%), Gaps = 6/175 (3%)

Query: 16 TVLTERLR LQPVELTNVND FLEFSSDSETVFYMQRYKANTVEEAQVVL A---NVC MKSPL 72

T+ TERL L+ +EL + + ++ SD E YM V +A+ ++ ++ ++

Sbjct: 3 TLYTERLTLRKMELEDADVL CQYWS DPEVTKYMNITPFTDVSQARDMIQMINDLSLEGQA 62

Query: 73 GIYAMIEKESQKMIGIIELEIRDEF S--AEFGYILNKNYNGKGYMTEACSKLMSIGFEHL 130

+++I KE+ ++IG + D+ + AE GY L +N+ GKG+ +EA KL+ GF L

Sbjct: 63 NRFSIIIVKETDEVIGTCGFNMIDQENGRAEIGYDLGRNHWGKGFASEAVQKLIDYGF TSL 122

Query: 131 DLERIYARFDINNKKSGNVMERIGMKKEGELRHLAKNPKGEWKTRAYYSILKEEY 185

+L RI A+ + N S ++ + +KEG LR K KG +S+LK EY

Sbjct: 123 NLNRIEAKVEPENTPSIKLLNSLSFQKEGLLRDYEK-AKGR LIDVYMFSLKREY 176

-278-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 209

A DNA sequence (GBSx0223) was identified in *S.agalactiae* <SEQ ID 669> which encodes the amino acid sequence <SEQ ID 670>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.5180(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA87001 GB:Z46902 unknown [Saccharomyces cerevisiae]
Identities = 105/224 (46%), Positives = 148/224 (65%), Gaps = 3/224 (1%)

Query: 1  MGDVVENFTEGKNPKIDTLNGKTVRIEKINPD-HFEDLFQVYGELSTEDSLTYISFSKFN 59
      +G VE +T  P+  L G T R+E ++ + H  +LF Y E  +  TY+  F
Sbjct: 11 VGADVEGWTTTRAFPEKVVVLKGNTCRLEPLDRERHGSSELFSAyseag-QKLWTYLPAGPFT 69

Query: 60  SKNEFDVFFQTLLKSEDPYYLAIVDNNNGKVLGTFSLMRIDTKNRVVMGWWVYSSKLKQ 119
      + E+  F + L +++D  AI++  T + +GT  L+RID N  +E+G+VV+S +L++
Sbjct: 70  NLEEYLEFIKELNETKDTVPFAIINKETERAVGTLCLIRIDEANGSLEVGYVVFSPQLQK 129

Query: 120 TRIATEAQYLVMKYVFEEELCYRRYEWKCDSLNAPSNNNSAKRLGFTFECTFRQAVVYKGRN 179
      T IATEAQ+L+MKYVF++L YRRYEWKCDSLN PS  +A RLGF +EGTFRQ VVYKGR
Sbjct: 130 TIATEAQFLLMKYVFDDLQYRRYEWKCDSLNGPSRRAMRLGFKYEGTFRQVVVYKGR 189

Query: 180 RDTNWYSILDKEWPEKKTRFEKWLDSDNFAVNGYQIRSLSSIEQ 223
      RDT W+SI+DKEW  + FE+WLD +NF  NG Q R +++I +
Sbjct: 190 RDTQWFSIIDKEWLRIRKTFEEWLDKTNFE-NGKQKRGIAAIRE 232
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 210

A DNA sequence (GBSx0224) was identified in *S.agalactiae* <SEQ ID 671> which encodes the amino acid sequence <SEQ ID 672>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood =-12.15    Transmembrane    25 - 41 ( 20 - 49)

----- Final Results -----
bacterial membrane --- Certainty=0.5861(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8519> and protein <SEQ ID 8520> were also identified. Analysis of this protein sequence reveals the following:



-279-

Lipop: Possible site: -1 Crend: 10

McG: Discrim Score: -3.31

GvH: Signal Score (-7.5): -4.44

Possible site: 39

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

ALOM program count: 1 value: -12.15 threshold: 0.0

INTEGRAL Likelihood = -12.15 Transmembrane 25 - 41 ( 20 - 49)

PERIPHERAL Likelihood = 11.94 59

modified ALOM score: 2.93

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.5861(Affirmative) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial cytoplasm --- Certainty=0.0000(Not Clear) &lt; succ&gt;

SEQ ID 672 (GBS43) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 5 (lane 4; MW 34kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 9; MW 58kDa) and in Figure 15 (lane 4; MW 59kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 211**

A DNA sequence (GBSx0225) was identified in *S.agalactiae* <SEQ ID 673> which encodes the amino acid sequence <SEQ ID 674>. Analysis of this protein sequence reveals the following:

Possible site: 32

&gt;&gt;&gt; May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial cytoplasm --- Certainty=0.0000(Not Clear) &lt; succ&gt;

A related GBS nucleic acid sequence <SEQ ID 9519> which encodes amino acid sequence <SEQ ID 9520> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 212**

A DNA sequence (GBSx0226) was identified in *S.agalactiae* <SEQ ID 675> which encodes the amino acid sequence <SEQ ID 676>. Analysis of this protein sequence reveals the following:

Possible site: 44

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -1.54 Transmembrane 165 - 181 ( 164 - 181)

INTEGRAL Likelihood = -0.85 Transmembrane 67 - 83 ( 67 - 84)

----- Final Results -----

bacterial membrane --- Certainty=0.1617(Affirmative) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial cytoplasm --- Certainty=0.0000(Not Clear) &lt; succ&gt;

-280-

The protein has homology with the following sequences in the GENPEPT database:

```

5  >GP:CAA82211 GB:Z28353 similar to a B.subtilis gene (GB:
    BACHEMEHY_5) [Clostridium pasteurianum]
    Identities = 40/185 (21%), Positives = 87/185 (46%), Gaps = 6/185 (3%)

    Query: 18  MPKGKQKVILSAIELFASQGFHGTSTAQLAKNAEVSQATIIKYFETKDKLLVFILELIVQ 77
              M K K + SAI++F++ G++G + ++A NA V++ T+Y +F++K+++ +I+E V
    10  Sbjct: 1   MNKTKDNIFYSAIKVFSNNGYNGATMDEIASNAGVAKGTLYYHFKSKEEIFKYIIEEGVN 60

    Query: 78  TIGRPFFTELSTFSTKEELIHFFVQDRPKFIEKNNDLIKILMQELLINSETSTIFTKLIN 137
              +      T E + + + I KN D K++ +L      ++
    15  Sbjct: 61  LMKNEIDEATDKETALEKLVKAVCRVQLNLIYKNRDFKVIASQLWGKELRQLELRDIMR 120

    Query: 138 STDPNITKIFNCLSEGNL---NKMEILRAVIGQFITFFIQLY-ILNIKPENLEELKQI 193
              + +I + E S+ N + + A +G + + LY ++N + +N+ ++ +
    20  Sbjct: 121 NYVVHIEEFVKDAMEAGSIKGNLFLVAYAFGLTLC--VSLYEVINAENDNINNTIENL 178

    Query: 194 EKQIL 198
              IL
    25  Sbjct: 179 MNYIL 183

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 213

A DNA sequence (GBSx0227) was identified in *S.agalactiae* <SEQ ID 677> which encodes the amino acid sequence <SEQ ID 678>. Analysis of this protein sequence reveals the following:

```

30  Possible site: 24
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2389(Affirmative) < succ>
    35  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 214

A DNA sequence (GBSx0228) was identified in *S.agalactiae* <SEQ ID 679> which encodes the amino acid sequence <SEQ ID 680>. Analysis of this protein sequence reveals the following:

```

45  Possible site: 43
    >>> Seems to have no N-terminal signal sequence

    INTEGRAL    Likelihood =-13.32    Transmembrane  341 - 357 ( 333 - 361)
    INTEGRAL    Likelihood =-10.93    Transmembrane  253 - 269 ( 238 - 277)
    INTEGRAL    Likelihood =-10.77    Transmembrane  172 - 188 ( 166 - 196)
    INTEGRAL    Likelihood = -8.01     Transmembrane  225 - 241 ( 215 - 251)
    50  INTEGRAL    Likelihood = -7.01     Transmembrane   21 - 37 ( 18 - 42)
    INTEGRAL    Likelihood = -2.66     Transmembrane  285 - 301 ( 283 - 301)

    ----- Final Results -----

```

-281-

bacterial membrane --- Certainty=0.6328(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB42664 GB:AL049819 putative integral membrane protein  
 [Streptomyces coelicolor A3(2)]  
 Identities = 60/156 (38%), Positives = 101/156 (64%), Gaps = 1/156 (0%)

10 Query: 176 LMGFMVFFVFLISGMALLKERTSGTLDRLLATPVKRSDIVFGYMLSYGILAIQTIVIV 235  
 L+G +FL++ +A L+ERTSGTL+RLLA P+ + D++ GY L++G LAI+Q+ +  
 Sbjct: 77 LLGIFPLITMFLVTSIATLRERTSGTLERLLAMPLGKGDLIAGYALAFGALAIVQSALAT 136

15 Query: 236 LSTIWLLDIQVVGSIIFSIIIVNFILALVALSLGILMSTLAKSEFQMMQFIPLIIMPQLFF 295  
 +W L + V GS + +++V + AL+ +LG+ +S A SEFQ +QF+P +I PQL  
 Sbjct: 137 GLAVWFLGLDVTGSPWLLLLVALLDALLGTALGLFVSAFAASEFQAVQFMPAVIFPQLLL 196

Query: 296 SGII-PLENMAWAQTVGKILPLSYSGDALTKIIMY 330  
 G+ P +NM + V +LP+SY+ D + +++ +

20 Sbjct: 197 CGLFTPRDNMHPALEAVSDVLPMSYAVDGMNEVLRH 232

There is also homology to a DNA sequence which was identified in *S.pyogenes* <SEQ ID 681> which encodes the amino acid sequence <SEQ ID 682>. Analysis of this protein sequence reveals the following:

Possible site: 39

25 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -11.41	Transmembrane	263 - 279 ( 246 - 284)
INTEGRAL	Likelihood = -7.70	Transmembrane	231 - 247 ( 224 - 258)
INTEGRAL	Likelihood = -4.99	Transmembrane	20 - 36 ( 18 - 39)
INTEGRAL	Likelihood = -3.72	Transmembrane	349 - 365 ( 345 - 368)
INTEGRAL	Likelihood = -3.45	Transmembrane	187 - 203 ( 182 - 204)

30 ----- Final Results -----

bacterial membrane --- Certainty=0.5564(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 35 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB12662 GB:Z99108 similar to ABC transporter (ATP-binding protein) [Bacillus subtilis]

40 Identities = 92/369 (24%), Positives = 180/369 (47%), Gaps = 25/369 (6%)

Query: 12 IKRKTSYVTFFLMPILTLLALSLSFSNNQAKIGILDKDNSQISKQFIAQLKQNKKYD 71  
 I +K +Y+ F P+L T + S+ N+++ ++ I+D+D++ +S+ +I QLK +  
 Sbjct: 15 IFKKPQNYLIMFAAPLLLTFFVFGSMLSGNDDKVRLLAIVDQDDTILSQHYIRQLKAHDDMY 74

45 Query: 72 IFTKIKKEHIDHYLDQKSLEAVLTIDKGFSDKVLQGKSQKLNIRSIANSEITEWVKAQTN 131  
 +F + + L+ K + ++ I + F ++ +GK +L R VK  
 Sbjct: 75 VFENMSESKASEKLKQKKIAGIIVISRSFQTQLEKKGHPHIFRHGPELSEAPMVKQYAE 134

50 Query: 132 YLENYNIIGDVALGNEDTFNR-----ILQKNQQLNYDVKQVTLTDRSRKAVSST 182  
 L NI A T +K++ + V + TL+D+ S T  
 Sbjct: 135 SALATLNIQVTAAKTASQTAGENWKAAYKTVFAKKHEDIVPAVTRQTLSDKKEGAEASDT 194

55 Query: 183 TT---GFLILMLGSTSVIYSGILADKSSQLYHRLMLSNLSRFR---YMLSYVCVGFVA 235  
 + GF ++ ++ + IL + + ++ RL+ +++SR Y+LS+ +G++  
 Sbjct: 195 ASRAAGFSILFVMLTMMGAAGTILEARKNGVWSRLLTASVSRAEIGAGYVLSFFVIGWV 254

Query: 236 FTIQIVIMLSLLKVFNISFFVPTSLLLIIFFLFSLAIGFLLIGAITQNSQQSSQLANL 295  
 F I ++LS +F I++ P ++++++ LF L +G GL+I A + +Q NL  
 60 Sbjct: 255 FGI---LLLSTHWLFGINWGNPAAVIVLVS-LFLLTVVGIGLMIAANVRTPEQQLAFGNL 310

Query: 296 IVMPTSMLAGCLWPLSITPSYMQAIGKLLPQNWVLSAIA-IFQSGGTLSQAWPYLLALMG 354  
 V+ T M++G WP+ I P +MQ+I + LPQ W +S + I +G ++ +L + G  
 Sbjct: 311 FVIATCMVSGMYWPIIDIEPKFMQSAEFLPQKWMASGLTEIIANGARVTD----ILGICG 366

-282-

Query: 355 TALALISFS 363  
 LA + +  
 Sbjct: 367 ILLAFAAIT 375

5

An alignment of the GAS and GBS proteins is shown below:

Identities = 92/375 (24%), Positives = 164/375 (43%), Gaps = 66/375 (17%)

10

Query: 11 IKELF----RDKRTLAMMFLAPILIMFLMNMVFSANSNTKVKIGTINVNTKVVSNLDNIK 66  
 IK LF R K + FL PIL L+ + S ++N + KIG ++ + +S  
 Sbjct: 5 IKTLFVKIKRKKTSYVTFFLMPILTT-LLALSLSFSNNQAKIGILDKDNSQISK----- 58

15

Query: 67 HIQVRSFKFNSSAKKALKSNKIDALISEDNKSYTVFYANTDSSKTTLT-RQAFKTAVENTM 125  
 +F + LK NK + ++ K + Y S + LT + F V  
 Sbjct: 59 -----QFIAQ---LKQNKYDIFTKIKKEHIDHYLQDKSLEAVLTIDKGFSDKVLQ 107

20

Query: 126 NSKELISQVKILANKNPKLAQSLQTRSKYIKEKYN-----GNKNT-----GF 168  
 S++L I + N ++ + ++ ++ Y+ E YN GN++T +  
 Sbjct: 108 KSQKL---NIRSIANSEITEWVKQTNVLLLENYNIIGDVALGNEDTFNRILQKNQQLNY 163

25

Query: 169 FAKMIPIL-----MGFMVFFVFLISGM--ALLKERTSGTLDRLLATPVKRSD 214  
 K + + GF++ + S + +L +++S RL+ + + R  
 Sbjct: 164 DVKQVTLTDRSRKAVSSTTTGFLILMLGSTSVIYSGILADKSSQLYHRLMLSNLRS-- 221

30

Query: 215 IVFGYMLSY---GILAIQITIVIVLSTIWLDDIQVVGSIIFSIIIVNFILALVALSLGILM 271  
 F YMSY G +A IVI+LS + + +I ++I+ F+ +L+A+ G+L+  
 Sbjct: 222 --FRYMLSYVCVGFVAFTTIQIVIMLSLLKVFNISFFVPTSLLLIIFFLSLLAIGFGLLI 279

35

Query: 272 STLAKSEFQMMQFIPLIIMPQLFFSGII-PLNMAWAQTVGKILPLSYSGDALTKIIMY 330  
 + ++ Q Q LI+MP +G + PL S+ Q +GK+LP ++ A+ I  
 Sbjct: 280 GAITQNSQQSSQLANLIVMPTSMLAGCLWPLSITPSYMQAIGKLLPQNWVLSAIA-IFQS 338

Query: 331 GQGLEPNVSSNLLVLL 345  
 G L LL L+

Sbjct: 339 GGTLQAWPYLLALM 353

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9081> which encodes the amino acid sequence <SEQ ID 9082>. Analysis of this protein sequence reveals the following:

40

Possible site: 38  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -12.52 Transmembrane 21 - 37 ( 17 - 43)  
 INTEGRAL Likelihood = -10.30 Transmembrane 351 - 367 ( 346 - 371)  
 INTEGRAL Likelihood = -5.36 Transmembrane 262 - 278 ( 260 - 285)  
 INTEGRAL Likelihood = -2.60 Transmembrane 288 - 304 ( 288 - 305)  
 INTEGRAL Likelihood = -1.81 Transmembrane 229 - 245 ( 229 - 246)

45

----- Final Results -----

50

bacterial membrane --- Certainty=0.6010(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS sequences follows:

Score = 62.5 bits (149), Expect = 9e-12

Identities = 72/382 (18%), Positives = 166/382 (42%), Gaps = 32/382 (8%)

55

Query: 1 MVLFHLLIKESLQIFRNRTALLMMVIFPILMIVILSFAPKSSFNATTPVKLTIRYQLEG 60  
 M + + +K ++FR++ L MM + PIL++ +++ F ++ NT + + + +  
 Sbjct: 1 MRIIAITEKVIKELFRDKRTLAMMFLAPILIMFLMNMVFSANSNTKVKIGTINVNTKVV 60

60

Query: 61 EKTDYQKNFLAFLKVLNQLHLETKPSNSLEKDRQVSEGALTAVLEVKKNTIKVITNN 120  
 L+ H++ + ++ + +A++ + N++ V N  
 Sbjct: 61 N-----LDNIKHQVRSFKFNSSAKKALKSNKIDALIS-EDNKSYTVFYAN 105

Query: 121 INQQNADLINMLVKNYVDNAKTYDSIAALY-----PQQLNHIRKRSVDYVKVSSIQTSK 174

-283-

```

      +      L      K  V+  + + I+ +      P+  ++ RS  Y+K  + +
Sbjct: 106 TDSSKTTLTRQAFKTAVENTMNSKELISQVKILANKNPKLAQSLQTRS-KYIKE---KYN 161

Query: 175 GMTSADYYA----ISMFTMITFYSMMSAMNLVLSDRQQRITNRIHLTGVSPSFLVFGKLI 230
      G  +  ++A  I M  M+ F+  + +  +L +R      +R+  T V  S  +VFG ++
Sbjct: 162 GNKNTGFFAKMIPILMGFMVFFFVFLISGMALLKERTSGTLDRLLATPVKRSDIVFGYML 221

Query: 231 GAMLATTVQLSLLYIFTRFVLRVNWGTNEWMLIGITASLVYLSVAIGIGLGISIKNEAFL 290
      +  +Q  ++ + T ++L +      + + +I +  L  +++++GI +  K+E  +
Sbjct: 222 SYGILAIITQIVIVLSTIWLDDIQVVGSI FSVIIVNFILALVALSLGILMSTLAKSEFQM 281

Query: 291 TVASNTIIPFAFLGGSYVPLTTLHSSIINQLSNISPIKWVNDLSFYLI FGGQYNP-IPV 349
      II  F  G  +PL  + +S      +  I P+ +  D+L  +I  GQ  P  +
Sbjct: 282 MQFIPLIIMPQLFFSG-IIPLENM-ASWAQTVGKILPLSYSGDALTKIIMYGQGLPNVSS 339

Query: 350 TLIVNISIGTIFIILALIGMRK 371
      L+V +      I  I  + G+++
Sbjct: 340 NLLVLLLLFLIILTIANIFGLKR 361

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 215

A DNA sequence (GBSx0229) was identified in *S. agalactiae* <SEQ ID 683> which encodes the amino acid sequence <SEQ ID 684>. This protein is predicted to be CG1718 gene product (b0794). Analysis of this protein sequence reveals the following:

```

Possible site: 61
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -1.17      Transmembrane 118 - 134 ( 117 - 134)

----- Final Results -----
      bacterial membrane --- Certainty=0.1468(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8521> which encodes amino acid sequence <SEQ ID 8522> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1      Crend: 8
McG: Discrim Score:      -10.96
GvH: Signal Score (-7.5): -4.84
      Possible site: 15
>>> Seems to have no N-terminal signal sequence
ALOM program count: 1 value: -1.17 threshold: 0.0
      INTEGRAL      Likelihood = -1.17      Transmembrane 142 - 158 ( 141 - 158)
      PERIPHERAL Likelihood = 4.98      197
modified ALOM score: 0.73

*** Reasoning Step: 3

----- Final Results -----
      bacterial membrane --- Certainty=0.1468(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAF50837 GB:AE003568 CG1718 gene product [Drosophila melanogaster]
Identities = 80/204 (39%), Positives = 123/204 (60%), Gaps = 3/204 (1%)

Query: 7      EIIGLIGPSGAGKSTLIKTMLGMEKADKGTALV--LDTQMPDRNILNQIGYMAQSDALYE 64
      E  GL+G +GAGK+T  K M G E+  G A V L  +  +I  IGY  Q DAL  +

```

-284-

Sbjct: 1394 ECFGLGVNGAGKTTTFKMMTGDERISSGAAYVQGLSLESNMNSTYKMGYCPQFDALLD 1453

Query: 65 SLTGLENLLFFGKMKGIQKTELKQQITHISKVVDLENQLDKFVSGYSGGMKRRLSLAIAL 124  
 LTG E L F ++G+Q++ ++Q ++K +DK YSGG KR+LS AIA+

5 Sbjct: 1454 DLTGREVLRIFCMLRGVQESRIRQLSEDLAKSFGFMKHIDKQTHAYSGGNKRKLSTAIIV 1513

Query: 125 LGNPTVLILDEPTVIGIDPSLRRKIWQELINIKDEGHSIFITTHVMDEAE-LTSKVALLR 183  
 +G+P+V+ LDEPT G+DP+ RR++W + I+D G SI +T+H M+E E L +++A+++

10 Sbjct: 1514 IGSPSVIYLDEPTTGMDDPAARRQLWNVCRIRDSGKSIVLTSHSMEECEALCTRLAIMVN 1573

Query: 184 GNIIAFDTPHLKKQFNVSTIEEV 207  
 G + HLK +F+ I ++

Sbjct: 1574 GEFKICIGSTQHLKNKFSKGLILKI 1597

15 Identities = 73/216 (33%), Positives = 128/216 (58%), Gaps = 9/216 (4%)

Query: 1 MEVFKGEIIGLIGPSGAGKSTLIKTM LGMEKADKGTALV--LDTQMPDRNINLQIGYMAQ 58  
 M +F+ EI L+G +GAGK+T I + GM GTA++ D + +G Q

Sbjct: 536 MNMFEDITVLLGHNGAGKTTTISMLTGMFPPTSGTAIINGSDIRTNIEGARMSLGICPQ 595

20 Query: 59 SDALYESLTGLENLLFFGKMKGIQKTELKQQITHISKVVDLENQLDKFVSGYSGGMKRRL 118  
 + L++ ++ ++ FF +MKG++ ++Q++ K+++LE++ + S SSGMKR+L

Sbjct: 596 HNVLFDEMSVSNHIRFFSRMKGLRGKAVEQEVAKYLMIELEDKANVASSKLSGGMKRKL 655

25 Query: 119 SLAIALLGNPTVLILDEPTVIGIDPSLRRKIWQELINIKDEGHSIFITTHVMDEAE-LTSK 177  
 S+ AL G+ V++ DEP+ G+DPS RR++W +L+ + G ++ +TTH MDEA+ L +

Sbjct: 656 SVCCALCGDTKVVLCDPESSGMDPSARRQLW-DLLQOEKVGRITLLTTHFMDEADVIGDR 714

Query: 178 VALLLRGNIIAFDTPHLKKQFN-----VSTIEEVF 208  
 +A++ G + T LKKQ+ VS ++ +F

30 Sbjct: 715 IAIMCDGELKCQGTSTFLLKKQYGSYRLVSGVQNLF 750

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 685> which encodes the amino acid sequence <SEQ ID 686>. Analysis of this protein sequence reveals the following:

Possible site: 59

35 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.43 Transmembrane 49 - 65 ( 49 - 65)

----- Final Results -----

40 bacterial membrane --- Certainty=0.1171(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

45 >GP:CAB12660 GB:Z99108 similar to ABC transporter (ATP-binding protein) [Bacillus subtilis]

Identities = 151/316 (47%), Positives = 202/316 (63%), Gaps = 18/316 (5%)

Query: 4 VQLTNVVKSYKNGKKA-VNDVSLSIEAGNIYGLLGPNGAGKSTLINLILGLIPLSSGKIT 62  
 +Q N+ K+Y GKK V +S S++ G +GLLGPNGAGKST I++I GL+P SG IT

50 Sbjct: 2 LQAENIKKAY--GKKTIVKGISFSLKKGESFLLGPNGAGKSTTISMISGLVPHDSGNIT 59

Query: 63 VLGQS-QKTIRKISSQIGYVPQDIIVPDLTAYENVELFGSLYGLKGAQLKKQVLKSLEF 121  
 V G K K +IG VPQ+IA+YP LTA+EN+ +G +YGL + KK+ + LE+

55 Sbjct: 60 VGGYVIGKETAKAKQKIGIVPQEIALLYPTLTAHENLMFWGKMYGLTHDEAKKRAAEVLEY 119

Query: 122 VGLHSQAKQFPSPQFSGGMKRRLNIACALVHSPKLIIFDEPTVIGIDPQSRNHILESIRLLN 181  
 VGL +AK FSGMKRR+NI AL+H P+L+I DEPTVIGIDPQSRNHILE+++ LN

60 Sbjct: 120 VGLTERAKDKIETFSGGMKRRLINIGALMHKPELLIMDEPTVIGIDPQSRNHILETVKQLN 179

Query: 182 KEGATVIYTHYMEVEEALCDYIFIMDHGQVIEEGPKFELEKRYVANLANQIIVTLTDSR 241  
 + G TVIYT+HYMEVEE LCD I I+D G++I G K +L R + Q+ V+ +

65 Sbjct: 180 ETGMTVIYTSHYMEVEEFLCDRIGIIDQGEMIAIGTKTDLCSRLGGDTIIQLTVSGINEA 239

Query: 242 HL-----ELADKPDWSLIEDGEKMLKIDNSD-----MTSVVHQLTQANITFSEIRHNHL 291  
 L LA D ++ E L LKID S +TS++ + T +I ++

-285-

Sbjct: 240 FLVAIRSLAHVNDVTVHE----LELKIDISAAHEKVVTSLLAEATAHHINLLSLQVQEP 295

Query: 292 NLEEIFLHLTGKKLRD 307

NLE +FL+LTG+ LRD

5 Sbjct: 296 NLERLFLNLTGRTLRLD 311

An alignment of the GAS and GBS proteins is shown below:

Identities = 81/211 (38%), Positives = 125/211 (58%), Gaps = 2/211 (0%)

10 Query: 1 MEVFKGEIIGLIGPSGAGKSTLIKIMLGMEKADKGTALVL-DTQMPDRNINLQIGYMAQS 59  
+ + G I GL+GP+GAGKSTLI +LG+ G VL +Q R I +QIGY+ Q  
Sbjct: 25 LSIEAGNIYGLLGPNGAGKSTLINLILGLIPLSSGKITVLGQSQKTIRKISSQIGYVPQD 84

15 Query: 60 DALYESLTGLENLLFFGKMKGIQKTELKQQITHISKVVDLENQLDKFVSGYSGGMKRRLS 119  
A+Y LT EN+ FG + G++ +LK+Q+ + V L +Q +F S +SGGMKRRL+  
Sbjct: 85 IAVYPDLTAYENVELFGSLYGLKGAQLKKQVLKSLEFVGLHSQAKQFPSQFSGGMKRRLN 144

20 Query: 120 LAIALLGNPVTVLILDEPTVGIDPSLRKIWQELINIKDEGHSIFITTHVMDEAE-LTSKV 178  
+A AL+ +P ++I DEPTVGIDP R I + + + EG ++ TTH M+E E L +  
Sbjct: 145 IACALVHSPKLIIFDEPTVGIDPQSRNHILESTIRLLNKEGATVIYTTTHYMEEVEALCDYI 204

25 Query: 179 ALLLRGNIIAFDTPHLHLKKQFNVSTIEEVFL 209  
++ G +I L+K++ + ++ +  
Sbjct: 205 FIMDHGQVIEEGPKFELEKRYVANLANQIIV 235

SEQ ID 8522 (GBS391) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 74 (lane 7; MW 30kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 83 (lane 4; MW 55kDa).

GBS391-GST was purified as shown in Figure 217, lane 3.

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 216

A DNA sequence (GBSx0230) was identified in *S.agalactiae* <SEQ ID 687> which encodes the amino acid sequence <SEQ ID 688>. Analysis of this protein sequence reveals the following:

35 Possible site: 13  
>>> Seems to have no N-terminal signal sequence  
  
----- Final Results -----  
40 bacterial cytoplasm --- Certainty=0.6732(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 217

50 A repeated DNA sequence (GBSx0231) was identified in *S.agalactiae* <SEQ ID 689> which encodes the amino acid sequence <SEQ ID 690>. This protein is predicted to be ISL2 protein. Analysis of this protein sequence reveals the following:

-286-

Possible site: 58

&gt;&gt;&gt; Seems to have an uncleavable N-term signal seq

----- Final Results -----

5                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10       >GP:CAC18596 GB:AJ278419 IS1381 transposase [Streptococcus pneumoniae]  
           Identities = 111/129 (86%), Positives = 117/129 (90%)

Query: 1   MKAQAIVTSQGRIVSLDI VNYCHDMKLFKMSRRNIGQAAKILADSGYQGIMKMYSAQT 60  
           MK QAIVTSQGRIVSLDI VNYCHDMKLFKMSRRNIGQA KILADSGYQG+MK+Y QAQT

15       Sbjct: 1   MKTQAIVTSQGRIVSLDITVNYCHDMKLFKMSRRNIGQAGKILADSGYQGLMKIYPQAQT 60

Query: 61   PRKSSKLKPLTLEDKTYNHTLSKERIKVENIFAKVKTFKIFSTTYRNRKRFGRLRMNLIA 120  
           RKSSKLKPLT+EDK NH LSKER KVENIFAKVKTFK+FASTTYR+ RKRFGRLRMNL A

20       Sbjct: 61   SRKSSKLKPLTVEDKACNHLSKERSKVENIFAKVKTFKMFSTTYRSHRKRFGRLRMNLSA 120

Query: 121   GMINRELGF 129  
           G+IN ELGF

          Sbjct: 121   GIINHELGF 129

25   No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 218**

30   A repeated DNA sequence (GBSx0232) was identified in *S.agalactiae* <SEQ ID 691> which encodes the amino acid sequence <SEQ ID 692>. This protein is predicted to be ISL2 protein. Analysis of this protein sequence reveals the following:

Possible site: 41

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

35                   bacterial cytoplasm --- Certainty=0.3996(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40   The protein has homology with the following sequences in the GENPEPT database:

          >GP:CAC18595 GB:AJ278419 IS1381 transposase [Streptococcus pneumoniae]  
           Identities = 110/125 (88%), Positives = 119/125 (95%)

45       Query: 1   MNYEASKQLTDVRFKRLVGVQRTTFEEMLA VLKTAYQQRKHAKGGRTPKLSLEDLLMATLQ 60  
           MNYEASKQLTD RFKRLVGVQRTTFEEMLA VLKTAYQ KHAKGGR PKLSLEDLLMATLQ

          Sbjct: 1   MNYEASKQLTDARFKRLVGVQRTTFEEMLA VLKTAYQLKHAKGGRPKLSLEDLLMATLQ 60

Query: 61   YMREYRTYEQIAADFGIHESNLIRRSQWVESTLIQSGFTISKTHLSAEDTVIVDATEVKI 120  
           Y+REYRTYE+IAADFG+HESNL+RRSQWVE TL+QSG TIS+T LS+EDTV++DATEVKI

50       Sbjct: 61   YVREYRTYEEIAADFGVHESNLLRRSQWVEVTLVQSGVTISRTPLSSEDTVMIDATEVKI 120

Query: 121   NRPKK 125  
           NRPKK

          Sbjct: 121   NRPKK 125

55

No corresponding DNA sequence was identified in *S.pyogenes*.



Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 219

A DNA sequence (GBSx0233) was identified in *S.agalactiae* <SEQ ID 693> which encodes the amino acid sequence <SEQ ID 694>. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.40	Transmembrane	130 - 146 ( 123 - 156)
INTEGRAL	Likelihood = -7.86	Transmembrane	169 - 185 ( 167 - 191)
INTEGRAL	Likelihood = -6.90	Transmembrane	100 - 116 ( 95 - 118)
INTEGRAL	Likelihood = -5.52	Transmembrane	199 - 215 ( 189 - 216)

----- Final Results -----

bacterial membrane	---	Certainty=0.5161(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04126 GB:AP001508 unknown conserved protein in others  
[Bacillus halodurans]

Identities = 47/207 (22%), Positives = 95/207 (45%), Gaps = 14/207 (6%)

Query: 7 LQKENTLLEGRIDNSNNQTYTDMIVYLRGA-SISPYHQELIRNDIVNMLLEAQERQASLV 65  
L K+N + N + Y D+++Y+R A S S E + +++ LLEAQ + S  
Sbjct: 6 LIKDNNEKRKLLTEENLKVYEDLLLYIRLAHSKSEQETEELLTELLDHLLEAQAKGKSAK 65

Query: 66 SVFGEDRHDFFINQVIKSTPKISKKEE-TLQRWDLAILLLTIQMIIFLGGYLITEALQOSV 124  
+VFG++ + +++I PK+ KE L + L++ T+ ++F G Y + V  
Sbjct: 66 AVFGDNPKQYADEIIGEIPKMVTKERFGLFAYGLSMFFATV--LVFSGIYRMLRYVVFQV 123

Query: 125 PDLIPITLLDVLFAIFISIIAVKIADTIYATYNFDK----SKEKKYFFRYIFLILSLII 180  
+ + + A+ +I ++ IA ++ + + + K F +I + +I  
Sbjct: 124 GEAVSEVYVGT--ALITTIASIVIAWMFVVFVQYFRWSCFRTINKVFEFFILWLGGMIP 181

Query: 181 AYILIGKYYHLP----FINIPLWIYLI 203  
+ Y P I IP+++Y +  
Sbjct: 182 FALFFALLYFTPNVGRMIEIPVLYYFV 208

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 220

A DNA sequence (GBSx0234) was identified in *S.agalactiae* <SEQ ID 695> which encodes the amino acid sequence <SEQ ID 696>. This protein is predicted to be minor extracellular protease epr precursor (epr).

Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -10.72	Transmembrane	10 - 26 ( 5 - 33)
----------	---------------------	---------------	-------------------

----- Final Results -----

bacterial membrane	---	Certainty=0.5288(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

-288-

A related GBS nucleic acid sequence <SEQ ID 8523> which encodes amino acid sequence <SEQ ID 8524> was also identified. Analysis of this protein sequence reveals the following:

```

5      Lipop Possible site: -1   Crend: 8
      McG: Discrim Score:      12.11
      GvH: Signal Score (-7.5): -4.02
      Possible site: 29
      >>> Seems to have an uncleavable N-term signal seq
      ALOM program count: 1 value: -10.72 threshold: 0.0
10      INTEGRAL    Likelihood = -10.72   Transmembrane  8 - 24 ( 5 - 33)
      PERIPHERAL  Likelihood = 13.74      219
      modified ALOM score: 2.64

      *** Reasoning Step: 3

15      ----- Final Results -----
          bacterial membrane --- Certainty=0.5288(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

20      !GB:Z99123 extracellular serine protease [Bacillus s...

      >GP:CAB15866 GB:Z99123 extracellular serine protease [Bacillus subtilis]
      Identities = 44/150 (29%), Positives = 80/150 (53%), Gaps = 14/150 (9%)

25      Query: 37  QMDTVESSVNHVSDSQLTEAQDMLDKFEKKPSEKLLKDVELALNKLSSSSKKEALQKRFK 96
          ++D V+S N      + +A+D + K EK  +++ +  + A+NKL N + K+ LQKR
      Sbjct: 428  RLDKVQSYRN-----VKDAKDKVAKAEKYKTQQTVDTAQTAINKLPNGTDKKNLQKRRLD 481

30      Query: 97  KAKDKYLKDEADKKATKDATDLVEILEQAPSEENVLKAEAAVNKLTVKESKEALQKRIDT 156
          + K +Y+      A+K A D V  E++ + +V A++A+ KL      K +LQKR++
      Sbjct: 482  QVK-RYI-----ASKQAKDKVAKAEKSKKKTVDVSAQSAIGKLPASSEKTSLOKRLNK 533

      Query: 157  VKTQYGLIGNQTPSSSVAETTEQGTANPAS 186
          VK+      Q+ S++ ++T+  A  S
35      Sbjct: 534  VKSTNLKTAQQSVSAAEKKSTDANAAKAQS 563
      Identities = 39/124 (31%), Positives = 64/124 (51%), Gaps = 2/124 (1%)

      Query: 35  TTQMDTVESSVNHVSDSQLTEAQDMLDKFEKKPSEKLLKDVELALNKLSSSSKKEALQKR 94
          +++ +++ +N V + L AQ +  EKK ++      + A+N+L  K ALQKR
40      Sbjct: 521  SSEKTSLOKRLNKVKSTNLKTAQQSVSAAEKKSTDANAAKAQS AVNQLQAGKDKTALQKR 580

      Query: 95  FKKAKDKYLKDEADKKATKDATDLVEILEQAPSEENVLKAEAAVNKLTVKESKEALQKRI 154
          K K K  EA K T A  V+  E+  ++++  A++AVN+L  K LQKR+
      Sbjct: 581  LDKVKKKVAAAEAKKVETAKAK--VKKAEKDKTKKSKTSAQSAVNQLKASNEKTKLQKRL 638

45      Query: 155  DTVK 158
          + VK
      Sbjct: 639  NAVK 642

```

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 697> which encodes the amino acid sequence <SEQ ID 698>. Analysis of this protein sequence reveals the following:

```

      Possible site: 41
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -4.99   Transmembrane  24 - 40 ( 23 - 43)
55      ----- Final Results -----
          bacterial membrane --- Certainty=0.2996(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
60

```

The protein has homology with the following sequences in the databases:

```

      >GP:CAB15866 GB:Z99123 extracellular serine protease [Bacillus subtilis]

```

-289-

Identities = 43/130 (33%), Positives = 71/130 (54%), Gaps = 8/130 (6%)

Query: 41 GSHPTQDQKVA---KHSKSAASLLKKAVKAVNDADRLATAAAIQEAQKAVDKLAESSKKK 97  
 G P + +K + + +K ++ LK A ++V+ A++ +T A +AQ AV++L K  
 Sbjct: 516 GKLPASSEKTSIQRLNKKVSTNLKTAQQSVSAAEKKSTDANAQAQSAVNQLQAGKDKT 575

Query: 98 TLQEQNLN-----VAKAKQEQEDAATQAVKAAEETLNQNLKDIAQKAVNDLSNKGKKAALQ 152  
 LQ++L+ VA A+ ++ + A VK AE+ + K AQ AVN L +K LQ  
 Sbjct: 576 ALQKRLDKVKKVAAAEAKVETAKAKVKAEKDKTKKSKTSAQSAVNQLKASNEKTKLQ 635

Query: 153 SRLDAILPAK 162

RL+A+ P K

Sbjct: 636 KRLNAVKKPK 645

Identities = 31/105 (29%), Positives = 53/105 (49%), Gaps = 1/105 (0%)

Query: 54 SKSAASLLKKAVKAVNDADRLATAAAIQEAQKAVDKLAESSKKKTLQEQNLNVAKAKQEQE 113  
 +++ S A +AV A++ I +A++ + +L S K L ++L+ ++ + +  
 Sbjct: 380 AQATDSAYAAAEQAVKAEQTKAQIDINKARELISQLENSDAKTALHKRLDKVQSYRNVK 439

Query: 114 DAATQAVKAAEETLNQNLKDIAQKAVNDLSNKGKKAALQSRLDAI 158

DA + KA E+ Q D AQ A+N L N K LQ RLD +

Sbjct: 440 DAKDKVAKA-EKYKTQQTVDTAQTAINKLPNGTDKKNLQKRLDQV 483

An alignment of the GAS and GBS proteins is shown below:

Identities = 61/233 (26%), Positives = 115/233 (49%), Gaps = 13/233 (5%)

Query: 2 SMKIDKKELLALIASIILLIFASVTFFLFKDHGTTQMDTVESSVNHVSDSQLTEAQDMLD 61  
 SM +KE L + S++ + + +F H TQ + S + + S L +A ++  
 Sbjct: 12 SMTKSQKEALYWMLSVLTITLIGGSCIFGSHPTQDQKVAKHSKS--AASLLKKAVKAVN 69

Query: 62 KFEKPKSEKLLKDVELALNKLNSSSKKEALQKRFKAKDKYLKDEADKKATKDATDLVEI 121  
 ++ + +++ + A++KL+ SSK+ LQ++ AK K +++A AT V+  
 Sbjct: 70 DADRLATAAAIQEAQKAVDKLAESSKKKTLQEQNLNVAKAKQEQEDA-----ATQAVKA 122

Query: 122 LEQAPSEENVLKAEAAVNKLTVKESKEALQKRIDTVKTQYGLIGNQTPSSSVAETTEQGT 181  
 E+ ++ A+ AVN L+ K K ALQ R+D + +I ++ P S E T+  
 Sbjct: 123 AEETLNQNLKDIAQKAVNDLSNKGKKAALQSRLDAILPAKPII-DEFPRQS-GEITDNSY 180

Query: 182 ANPASQDTSSVYNQNVAPTYE-QPQANNTPTVPGVNNTPV-TPGTGTVPATNG 232

P D S + + +PT + +++ + VTP ++ P P T + P+ +G

Sbjct: 181 WTPFPGDVSDTYDNSQSPTLDPSESSASDVTQPQSHDPPIPPQTSSEPSDSG 233

SEQ ID 8524 (GBS278) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 6; MW 40kDa).

The GBS278-His fusion product was purified (Figure 206, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 305), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 221

A DNA sequence (GBSx0235) was identified in *S.agalactiae* <SEQ ID 699> which encodes the amino acid sequence <SEQ ID 700>. Analysis of this protein sequence reveals the following:

Possible site: 53

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1466(Affirmative) &lt; succ&gt;

bacterial membrane --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 222

A DNA sequence (GBSx0236) was identified in *S.agalactiae* <SEQ ID 701> which encodes the amino acid sequence <SEQ ID 702>. This protein is predicted to be N-acetylglucosamine-6-phosphate deacetylase (nagA). Analysis of this protein sequence reveals the following:

Possible site: 15  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.4607 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9297> which encodes amino acid sequence <SEQ ID 9298> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG21688 GB:AY007718 N-acetylglucosamine-6-phosphate deacetylase  
[Lactococcus lactis subsp. cremoris]  
Identities = 113/178 (63%), Positives = 135/178 (75%)

25 Query: 131 GIYFEGPYFTTEYKGAQNPIYMRNP NLEEFQWQKAAKGLITKIALAPEREGVEEFVSAI 190  
GI+FE GP+FTEE KGAQNP YMR+ + E WQ+AA G++ KI LAPEREG E+F+  
Sbjct: 1 GIFFEGPPFTTEKKGAQNPKYMRDAKMWELEDWQEAHGM LKKIGLAPEREGSEDFIRKA 60

30 Query: 191 TKQGVTV ALGHSNGTYKEAKKAVKAGASVWVHAYNGMRGLTHREPGMVGA VYNLPNTYAE 250  
T+ GV +ALGHSN TYK+A V+AGASVWVH +NGM G+TH+EPGMVGA+ N PNTYAE  
Sbjct: 61 TEGSVVIALGHSNATYKQAVAGVQAGASVWVHTFNGMSGMTHQEPMVGAILNTPNTYAE 120

35 Query: 251 LICDGHVDPVACDILMTQKGHNHVALITDCMAAGGAPDGDYMLGELPVVVSNGTARL 308  
LICDGHV P A +I++ KG +HV LITD M A G PDG YMLGE V V +G A L  
Sbjct: 121 LICDGHV RPEAAEIVVKMGADHVVLITDSMRAAGLPDGPYMLGEYEV EVRDGA A WL 178

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 703> which encodes the amino acid sequence <SEQ ID 704>. Analysis of this protein sequence reveals the following:

40 Possible site: 40  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.3114 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 227/300 (75%), Positives = 262/300 (86%)

50 Query: 9 MTKYIKADRFFYADHV KENGYLEIKDNHFGKWIENISGQEEILDYSGYQIAPGLVDTHIH 68  
MT Y+KAD F+Y V+ GYL + D FG+W E + +I+DY+GYQIAPGLVDTHIH  
Sbjct: 1 MTCYLKADCFYYPTEVRPAGYLSLHDGVFGWEVTEIVPADAQIIDYTG YQIAPGLVDTHIH 60

-291-

Query: 69 GFAGADVMDCDSEGLILMSAGLLSTGVTSFLPTTLTSDTKRLEEASKSVAAGKEQGAK 128  
 G+AGADVMD ++GI +MS GLL+TGVTSLPTTLTS ++LE+ S ++A+VA + +GAK  
 Sbjct: 61 GYAGADVMDNSAQGIHQMSSEGLLATGVTSFLPTTLTSTFEQLEKVS GTIASVADQVKGAK 120

5 Query: 129 IQGIYFEGPYFTEEYKGAQNPIYMRNPNLLEEFQWQKAAKGLITKIALAPEREGVEEFVS 188  
 IQGIYFEGPYFTEEYKGAQNP YM+ P LEEF WQKAAKGLI KIALAPER+CV+EFVS  
 Sbjct: 121 IQGIYFEGPYFTEEYKGAQNPSYMKTPRLEEFDAWQKAAKGLIKIALAPERDGVKEFVS 180

10 Query: 189 AITKQGVTVLGHSTNGTYKEAKKAVKAGASVWVHAYNGMRGLTHREPGMVGAVYNLPNTY 248  
 A+TKQGVTVLGHSTNGTY+EAK+AV+AGASVWVHAYNGMRGLTHREPGMVGAVYNLPNTY  
 Sbjct: 181 AVTKQGVTVLGHSTNGTYQEAKAVQAGASVWVHAYNGMRGLTHREPGMVGAVYNLPNTY 240

15 Query: 249 AELICDGHVDPVACDILMTQKGNHVALITDCMAAGGAPDGDYMLGELPVVVSNGTARL 308  
 AELICDGHV P+ACDILM QKGH+HVA+ITDCM AGG+PDGDY+LGE VVV+NGTARL  
 Sbjct: 241 AELICDGHVSPACDILMQQKGDHVMITDCMRAGGSPGDYLLGEFSVVVANGTARL 300

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 223

20 A DNA sequence (GBSx0237) was identified in *S.agalactiae* <SEQ ID 705> which encodes the amino acid sequence <SEQ ID 706>. Analysis of this protein sequence reveals the following:

Possible site: 25  
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3709(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 A related GBS nucleic acid sequence <SEQ ID 9307> which encodes amino acid sequence <SEQ ID 9308> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB16112 GB:Z99124 yyaQ [Bacillus subtilis]  
 Identities = 40/110 (36%), Positives = 62/110 (56%), Gaps = 12/110 (10%)

35 Query: 121 IAKTFEDSVDPYPAKHPQYASIRVSG--KWYALLFPLKMGKLENVPAQLSED---EVEVL 175  
 + + + S DYP+ K+P YAS R + KWY L+ + +P +L D E+++L  
 Sbjct: 11 VKEKYGTSPDYPEWKYPNYASLRHTSNKKWYGLIMNV-----LPEKLGLDGHGEIDIL 63

40 Query: 176 NIKVNPQDMEILLQKEGIYPSYHMSKKTWVSIVLNTLSDIEIFKLVSIDS 225  
 N+K P+ + L E I P YHM K+ W+SIVL+ T + EI+ L+ S  
 Sbjct: 64 NLKCPPEISDRLRNGENILPGYHMDKEHWISIVLERTDPEGEIYNLIEQS 113

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 707> which encodes the amino acid sequence <SEQ ID 708>. Analysis of this protein sequence reveals the following:

Possible site: 22  
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2541(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

55 Identities = 114/247 (46%), Positives = 169/247 (68%), Gaps = 1/247 (0%)  
 Query: 7 MSIESDFFRKRFIFSSLEEFQFIKSDQEIYICQTFMDNDFKAITITISLDGKIAGKVIDS 66

-292-

MS+ +D+F ++ I L +GF K D Y Y + FM+ +F+A + I G I +VID  
 Sbjct: 1 MSLATDYFSRQTFIVEKLMAYGFKEKRDNGYFYNERFMEGEFEAQLRIDEAGNIWDRVIDC 60

Query: 67 ALEEEYLPLRAANYNGSFVGEVRSAYMAILGDISDCKDLLFTKDQSNRLAEKIAKTFE 126  
 LEE+YLPL+ A + G++ G+VR+AY+ +L +S +C + F Q+NRLA+ I K +  
 Sbjct: 61 DLEEDYLPLQAAWQGTYYTGQVRAAYLELLERLSVACFEATPFQSMQANRLAKHITKEWS 120

Query: 127 DSVDYPPFAKHPQYASYRVSGKWYALLFPLKMGKLENVPAQLSEDEVEVLNIKVNPDMEI 186  
 D +DYPP KHP A+YRV GKWYA++F L KL+ +P +L EV+ +KVN+  
 Sbjct: 121 DPMDYPPFEKHPDLATYRVGGKWYAMIFSLADKLDQIPERLVGQTCVMTVKVNPKAFFQ 180

Query: 187 LLQKEGIYPSYHMSKKTWVSIVLDNTLSDIEIFKLVSRSRKLVSHNKSN-SEPEFWIIP 245  
 LLQ+EGIYP+YHMSK W+SI+LD+ ++D +++ LV+ SR+LV+ N SN + P++W+IP  
 Sbjct: 181 LLQKEGIYPAYHMSKKNWISILDDKLTDDKLTQSRQLVNPNGLSNPNPGPDYWVIP 240

Query: 246 ANPKFYD 252  
 AN K+YD  
 Sbjct: 241 ANLKYD 247

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 224

A DNA sequence (GBSx0238) was identified in *S.agalactiae* <SEQ ID 709> which encodes the amino acid sequence <SEQ ID 710>. This protein is predicted to be transposase for insertion sequence element is905.

Analysis of this protein sequence reveals the following:

Possible site: 61  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1824(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9601> which encodes amino acid sequence <SEQ ID 9602> was also identified.

A related GBS nucleic acid sequence <SEQ ID 9595> which encodes amino acid sequence <SEQ ID 9596> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA25167 GB:L20851 transposase [Lactococcus lactis]  
 Identities = 325/391 (83%), Positives = 365/391 (93%)

Query: 12 MTQFTTELLNFLAQKQDIDEFFRSSLTAMNDLLQVELSAFLGYEPYDKAGYNTGNSRNG 71  
 MTQFTTELLNFLAQKQDIDEFFR+SLETAMNDLLQ ELSAFLGYEPYDK GYN+GNSRNG  
 Sbjct: 1 MTQFTTELLNFLAQKQDIDEFFRSTLETAMNDLLQAELSAFLGYEPYDKVGYNSGNSRNG 60

Query: 72 AYTRRFETKYGVVNLIPDRNGEFSFALIPSYGRRDNHLEEMVIKLYRTGVTTREISDI 131  
 +Y+R+FEKYG V L IPRDRNG FSPAL+P+YGRRD+HLEEMVIKLY+TGVTTREISDI  
 Sbjct: 61 SYSRQFETKYGTVQLSIPDRNGNFSPALLPAYGRDDHLEEMVIKLYQGTGVTTREISDI 120

Query: 132 IERMYGHHYSPATVSNISKATQENVASFHERSLEANYTVLYLDGTYLPLRRGTVSKECIH 191  
 IERMYGHHYSPAT+SNISKATQENVA+FHRSLEANY+VL+LDGTYLPLRRGTVSKECIH  
 Sbjct: 121 IERMYGHHYSPATISNISKATQENVATFHRSLEANYSVLFLDGTYLPLRRGTVSKECIH 180

Query: 192 IALGVTSTYGHKAILGYDIAPNENNASWSDLLERFKGQGVQVSLVSDGFNGLDQLIQA 251  
 IALG+T G KA+LGY+IAPNENNASWS LL++ + QG+QVSLVV+DGF GL+Q+I QA  
 Sbjct: 181 IALGITPEGQKAVLGYEIAPNENNASWSTLLDKLQNGQIQVSLVVTDFGKLEQIISQA 240

Query: 252 FPMKQQRCLVHIGRNIASKVKRADRALILEQFKTIYRAINVEEAKQALDSFINWKPHY 311

-293-

+P+AKQQRCL+HI RN+ASKVKRADRA+ILEQFKTIYRA N+E A QAL++FI EWKP Y  
 Sbjct: 241 YPLAQQRCLIHISRNASKVKRADRAVILEQFKTIYRAENLEMAVQALENFIAEWKPKY 300

Query: 312 KKVIETLESIENTLLIFYEPHQIWGSIYSTNLIESLNKEIKRQTKKKVFPNEESLERYL 371  
 +KV+E+LE+ +NLL FY+FP+QIW SIYSTNLIESLNKEIKRQTKKKV+FPNEE+LERYL  
 Sbjct: 301 RKMESLENTDNLTFYQFPYQIWHSIYSTNLIESLNKEIKRQTKKKVLPNEEALERYL 360

Query: 372 VTLFSDYNFKQGQRIHKGFGQCTDTLESFLD 402  
 VTLF DYNFKQ QRIHKGFGQC DTLESFLD  
 Sbjct: 361 VTLFEDYNFKQSRIHKGFGQCADTLESFLD 391

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 711> which encodes the amino acid sequence <SEQ ID 712>. Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3054(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 111/128 (86%), Positives = 122/128 (94%)

Query: 12 MTQFTTELLNFLAQKQDIDIFFRSSLETAMNDLLQVELSAFLGYEPYDKAGYNTGNSRNG 71  
 MTQFTTELLNFLAQKQDIDIFFRSSLE AMNDLLQVELSAFLGYEPY+K GYNTGNSRNG  
 Sbjct: 1 MTQFTTELLNFLAQKQDIDIFFRSSLEIAMNDLLQVELSAFLGYEPYEKGYNTGNSRNG 60

Query: 72 AYTRRFETKYGVNLLIPDRNGEFSPALIPSYGRRDNHLEEMVIKLYRTGVTTREISDI 131  
 Y+R+FETKYG+VNL+IPDRNGEFSP L+PSY RR++HLEE+VIKLY+TGVTTREISDI  
 Sbjct: 61 TYSRQFETKYGLVNLIIPDRNGEFSPVLLPSYARREDHLEEIVIKLYQTGVTTREISDI 120

Query: 132 IERMYGHH 139  
 I+RMYG H  
 Sbjct: 121 IKRMYGDH 128

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 225

A DNA sequence (GBSx0239) was identified in *S.agalactiae* <SEQ ID 713> which encodes the amino acid sequence <SEQ ID 714>. Analysis of this protein sequence reveals the following:

Possible site: 32  
 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -12.42 Transmembrane 268 - 284 ( 260 - 286)  
 INTEGRAL Likelihood = -6.32 Transmembrane 232 - 248 ( 231 - 254)

----- Final Results -----

bacterial membrane --- Certainty=0.5967(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD40365 GB:AF036485 hypothetical protein [Plasmid pNZ4000]  
 Identities = 69/283 (24%), Positives = 133/283 (46%), Gaps = 9/283 (3%)

Query: 11 INVDDLSLQEERF-LPSELLAYARDENESS-FVRDIEGHLALVYQLLDTQGHVDDVRHVP 68  
 IN ++ + E+++ + +++ Y D +ES+ +V DI L L D +R++  
 Sbjct: 19 INAEERATLEDQYGIDEDIIEYVTDNDESTNYVVDINEDDQLFIPLAPYALDKDALRYIT 78

-294-

Query: 69 RVIPVTLFLKEDGLFVLANKHNINLVKKALNRV---EKVDSPKHLLLSLVTAFSKQYFDV 125  
 + P + L + LF N I V AL +V S +L + + +  
 Sbjct: 79 Q--PFGMLLHKGVLF--TF--NQSGIPEVNTALYSALDNPEVKSVDAFILETLETFVVVSFIPI 135

5 Query: 126 LDTISEERDKLINDLRKRPNKSNLARLANLQSGTVHLMGKQNFEMLTDLQNIQDKEN 185  
 I+++R+ L L ++ S+L L+ LQ L + N L L  
 Sbjct: 136 SRAITKKRNYLDKMLNRKTKNSDLVLSYLLQOTLTLFLSSAVQTNLSELDRLPKTHFGVGA 195

10 Query: 186 TRNEKMLQDAIIEARQLSNMCSLNSQVFQELS-SYNNVLSNNLNDNVTLTIISIGISI 244  
 +++ +D IE Q+ M + +QV + + N++ +NNLND + LTI S+ +++  
 Sbjct: 196 DQDKIDLFEDVQIEGEQVQRMFEIETQVVDRIHTLNSLANNLNDTMKFLTIWLSLTMV 255

Query: 245 IAMVTSFYGMNVKLPFDSVDAVWVLIILITITITIMLSIVMYI 287  
 +++ FYGMNVKLP + W+L + I+ ++ + + I++ +  
 15 Sbjct: 256 PTIISGFYGMNVKLPAGMQYAWMLTLGISVVLIVAMLIMLV 298

SEQ ID 714 (GBS422) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 7; MW 60kDa).

GBS422-GST was purified as shown in Figure 219, lane 12.

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 226

A DNA sequence (GBSx0240) was identified in *S.agalactiae* <SEQ ID 717> which encodes the amino acid sequence <SEQ ID 718>. Analysis of this protein sequence reveals the following:

25 Possible site: 45  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 30 bacterial cytoplasm --- Certainty=0.0783(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

35 >GP:CAB61731 GB:AL133220 putative oxidoreductase. [Streptomyces  
 coelicolor A3(2)]  
 Identities = 100/306 (32%), Positives = 152/306 (48%), Gaps = 3/306 (0%)

Query: 3 KVRYGCVSTAKVAPRFIEGVRLAGNGEVVAVSSRTLESAQAFANKYHLPKAYDKLEDMLA 62  
 KVR+G+++T +A RF + + EVVAV+SRT SA+ FA ++ +P+AY E +  
 40 Sbjct: 8 KVRWGILATGGMARFTADLVLDLPDAEVVAVASRTEASAKTFAERFGIPRAYGGWETLAR 67

Query: 63 DESIDVIYVATINQDHYKVAKAALLAGKHVLEKPFITLTYDQANELFALAESC�FLMEA 122  
 DE +DV+YVAT + H A L AG++VL EKPFTL +A EL ALA +FLMEA  
 45 Sbjct: 68 DEDVDVYVATPHSAHRTAAGLCLEAGRNVLCEKPFITLNAREAAELVALARENGVFLMEA 127

Query: 123 QKSVFIPMTQVIKLLASGEIGEVISISSTTAYPN-IDHVTWFRELELGGGTVHFMAPYA 181  
 P+ + +K+L+A G IGEV S+ + R+ GGG + + Y  
 Sbjct: 128 MMYCNPLVRRLEKELVADGATGEVRSIQADFLAGPFPAHRLRDPAQGGGALLDLGVYP 187

50 Query: 182 LSYLQYLFDATITHASGTATFPKGQSDSQSKLLQLSNGVLVDIFLTTRNLNPHEMIY 241  
 +S+ Q L T + A + D Q+ LL N L I + P+ I G  
 Sbjct: 188 VSFAQLLLGEP-TDVAARAVLSEEGVDLQTGALLSYGNDALASIHCSITGGTPNSASITG 246

Query: 242 TEGRLIIPH-FWKTTHAKLVRNDTSARTIQVDMVSDFEKEAYHVSQMILEGQRVSHIMTP 300  
 +EGR+ +P+ F+ H L R + + D + H ++ ++ R +P  
 55 Sbjct: 247 SEGRIDVPNGFFFPDHFVLRHTGRDPQEFRADPADGPRESLRHEAEVVMRALRAGETESP 306

Query: 301 QLTLSG 306  
 + L G



-295-

Sbjct: 307 LVPLDG 312

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## 5 Example 227

A DNA sequence (GBSx0241) was identified in *S.agalactiae* <SEQ ID 721> which encodes the amino acid sequence <SEQ ID 722>. This protein is predicted to be valyl-tRNA synthetase (valS). Analysis of this protein sequence reveals the following:

Possible site: 36

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.00 Transmembrane 794 - 810 ( 794 - 810)

----- Final Results -----

bacterial membrane --- Certainty=0.1001(Affirmative) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial cytoplasm --- Certainty=0.0000(Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database:

&gt;GP:AAA57558 GB:L08854 valyl-tRNA synthetase [Lactobacillus casei]

Identities = 543/881 (61%), Positives = 679/881 (76%), Gaps = 12/881 (1%)

Query: 5 LSPKYNPAEVEEGRYQTWLDQDVFKPSGDTEAKPYSIVIPPNVTGKHLGHAWDTTLQD 64

L+PKY+ VEEGRYQ WLD+DVFKPSGD +AKPYSIVIPPNVTGKHL+GHAWDTTLQD

Sbjct: 27 LAPKYDHKAVEEGRYQEWLDEDVFKPSGDKKAKPYSIVIPPNVTGKLHMGHAWDTTLQD 86

Query: 65 IIRQKRMQGFDTLWLPGMDHAGIATQAKVEERLREQGISRYDLGREKFLDKVWEWKDEY 124

I+IRQKR++GFDTLWLPGMDHAGIATQAKVE +LR++GISRYDLGREKF+ KVWEWKDE+

Sbjct: 87 IVIRQKRIEGFDTLWLPGMDHAGIATQAKVEAKLRKEGISRYDLGREKFVQKVWEWKDEF 146

Query: 125 AATIKSQWQKMGSLVDYSRERFTLDEGLSKAVRKVFVDLYNKGWIYRGEFIINWDPAART 184

A TI QW KMGLS+DYSRERFTLD+GL++AVR+VFVDLYN+G IYRGE+I+NWDPAART

Sbjct: 147 AKTIHQWAKMGLSLDYSRERFTLDEGLNQAARRVVFVDLYNQGLIYRGEYIVNWDPAART 206

Query: 185 ALSDIEVIHKDVEGAFYHMNYMLEDGSRALVATTRPETMFGDVAVAVNPEDARYKDLIG 244

ALSDIEVIHKD +GAFYH+ Y DGS +E+ATTRPETM GD AVAV+P D RYKD++G

Sbjct: 207 ALSDIEVIHKDDKGAFYHVKYPFADGSGYIEIATTRPETMMGDTAVAVHPGDERYKDMVG 266

Query: 245 QNVILPIINKPIPIVADEHADPEFGTGVVKITPAHDPNDFAVGQRHNLQVNVNDDGTM 304

+ILP+ N+ IPI+ D + DPEFGTG VKITPAHDPNDF VG RH+L ++N MNDDGTM

Sbjct: 267 TELILPLANRKIPIIEDAYVDPEFGTGAVKITPAHDPNDFQVGNRHLKRIINTMNDDGTM 326

Query: 305 NELADEFNMGDRFEARKAVVAKLESGLNVLKIKKTTTHSVGHSERTGVVVEPRLSTQWFWK 364

NE A ++ GMDRFEARKA+VA L+ G L+K++ HSVGHSERTGV VE RLSTQWFWK

Sbjct: 327 NENAGKYQGMDRFEARKAMVADLDKAGLLKVEPIVHSVGHSERTGVQVEARLSTQWFWK 386

Query: 365 MDQLAKNAI-ANQDTEKVEFYPPRFNDTFMSWMENVHDWVISRQLWWGHQIPAWYN-VN 422

M LA+ AI A Q+ + KV F P RF T++ WMEN+HDWVISRQLWWGHQIPAWYN

Sbjct: 387 MKPLAAEAAIKAQEPDKKVTTFVPERFEHTYQLWMENIHDWVISRQLWWGHQIPAWYNKQT 446

Query: 423 GEMYVGEDAPEG-DGWTQDEDVLDTWFSALWPFSTMGWPDTEAADFKRYFPTSTLVTGY 481

GE YVG +AP+ + W QD DVLDTWFSALWPFSTMGWP+T+A D+KRY+PT TLVTGY

Sbjct: 447 GETYVGMPEAPKDIEWQDPPDVLDTWFSALWPFSTMGWPNTDAPDYKRYPTDTLVTGY 506

Query: 482 DIIFFWVSRMIFQSLEFTGRQPFNSVLIHGLIRDEEGRKMSKSLGNGIDPMDVIEKYGAD 541

DII FWV+RMIFQ L FT ++PF LIHGL+RDE+GRKMSKSLGNGIDPMDVIEKYGAD

Sbjct: 507 DIIFFWVARMIFQGLHFTQRPFQYTLIHGLMRDEQGRKMSKSLGNGIDPMDVIEKYGAD 566

Query: 542 ALRWFLSNGSAPGQDVRFSYKMDASWNFINKIWNISRYILMNNEGLTLDQARENVEKVV 601

ALRWFL G+ PGQD RFSY++++A+WNFINKIWNISR+++MN L Q +

Sbjct: 567 ALRWFLITGNKPGQDTRFSYKQVEAAWNFINKIWNISRFVMMNLGDLDTPOQPD----- 620

-296-

Query: 602 NSQVGNVTDRLWILHNLNETVGKVTENFDKFEFGVAGHILYNFIWEEFANWYVELTKEVLY 661  
 +++D+W+ LNET+ +V + +FEFG G LYNF W A+WYVE++KEVLY  
 Sbjct: 621 -PSTFDLSDKWLFQQLNETIKQVMDLSARFEFGEMGRITLYNFTWNLADWYVEMSKEVLY 679

5 Query: 662 SDNEDEKVIITRSVLLYTLDQILRLHHPMPFVTEEIF--GQYAEGSIVLASYPQVNATFE 719  
 D+E K R L Y LDQILRLHHP+MPFV +++ + SIV ASYP N FE  
 Sbjct: 680 GDDEQAKAAKRVNLAYALDQILRLHHPMPFVHGKLLWALPHTGKSIVTASYPVANTAFE 739

10 Query: 720 NQTAHKGVESLKDILRSVRNSRAEVNVAAPSKPITILVKTSDSELESFFKDNSNYIKRFTN 779  
 N A ++++ LIR VR R E + ILVK +D L+ F+ N ++I RF N  
 Sbjct: 740 NADATSAMDAIIALIRGVRGIRKEAGAPLKTKVDILVKLTDPALKPIFEQNFDFIDRFVN 799

15 Query: 780 PETLEISSAIATPELAMSSVITGAEIFLPLADLINVEEELARLEKELAKWQKELDMVGKK 839  
 + + + +A P++A S+VITGA IF+PL +L++++EE A+L K+ K ++E+ + KK  
 Sbjct: 800 SKAFTVGTDAEPKMAGSAVITGATIFVPLNELIDLDEEKAKLTKDAKKLEQEIARIDKK 859

20 Query: 840 LSNERFVANAKPEVVQKEKDKQTDYQTKYDATIARIEEMKK 880  
 L+N+ F++ A VV +++ K++D++ + +T R+E++++  
 Sbjct: 860 LNNQGFLSKAPEAVVAEQRTKRSDFEDQLTSTKQRLQLQR 900

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 723> which encodes the amino acid sequence <SEQ ID 724>. Analysis of this protein sequence reveals the following:

Possible site: 61  
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.5062(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 782/878 (89%), Positives = 818/878 (93%)

35 Query: 4 ELSPKYNPAEVEEGRYQTWLDQDVFVKPSGDTEAKPYSIVIPPNVTGKLHLGHAWDTTLQ 63  
 ELSPKYNPAEVE GRYQ WLD DVFKPSGD +AKPYSIVIPPNVTGKLHLGHAWDTTLQ  
 Sbjct: 3 ELSPKYNPAEVEAGRYQKWLDADVFVKPSGDQKAKPYSIVIPPNVTGKLHLGHAWDTTLQ 62

40 Query: 64 DIIIRQKRMQGFDTLWLPGMDHAGIATQAKVEERLREQGISRYDLGREKFLDKVWEWKDE 123  
 DIIIRQKRMQGFDTLWLPGMDHAGIATQAKVEERLREQGISRYDLGR+KFLDKVWEWKDE  
 Sbjct: 63 DIIIRQKRMQGFDTLWLPGMDHAGIATQAKVEERLREQGISRYDLGRDKFLDKVWEWKDE 122

45 Query: 124 YAATIKSQWGKMGLSVDYSRERFTLDEGLSKAVRKVFVDLYNKGWIYRGEFIINWDPAAR 183  
 YA TIK QWGKMGLSVDYSRERFTLDEGLSKAVRKVFVDLY KGWYIRGEFIINWDPAAR  
 Sbjct: 123 YATTIKEQWGKMGLSVDYSRERFTLDEGLSKAVRKVFVDLYKKGWYIRGEFIINWDPAAR 182

50 Query: 184 TALSDIEVIHKDVEGAFYHMNYMLEDGSRALVATTRPETMFGDVAVAVNPEDARYKDLI 243  
 TALSDIEVIHKDVEGAFYHMNYMLEDGSRAL+VATTRPETMFGDVAVAVNPED RYKDLI  
 Sbjct: 183 TALSDIEVIHKDVEGAFYHMNYMLEDGSRALQVATTRPETMFGDVAVAVNPEDPRYKDLI 242

55 Query: 244 GQNVILPIINKPIPIVADEHADPEFGTG VVKITPAHDPNDFAVGQRHNLQVNVNMDDGT 303  
 G+NVLPI+NK IPIV DEHADPEFGTG VVKITPAHDPNDF VGQRHNLQVNVNMDDGT  
 Sbjct: 243 GKNVILPIVKNLPIVGDHADPEFGTG VVKITPAHDPNDFEVGQRHNLQVNVNMDDGT 302

60 Query: 304 MNELADEFNMGDRFEARKAVVAKLESLGNLVKIKKTTHSVGHSERTGVVVEPRLSTQWV 363  
 MNELA +F GMDRFEAR+A VAKLE LG LV I+K HSVGHSE+G VVEPRLSTQWV  
 Sbjct: 303 MNELAGDFAGMDRFEARQATVAKLEELGALVNIKRVHSVGHSE+GAVVEPRLSTQWV 362

65 Query: 364 KMDQLAKNAIANQDTEKVEFYPPRFNDTFMSWMENVHDWVISRQLWWGHQIPAWYNVNG 423  
 KMD+LAK A+ NQ+T+D+V+FYPFRFNDTF+ WMENVHDWVISRQLWWGHQIPAWYN G  
 Sbjct: 363 KMDELAKQAMDNQETDDRVDFYPPRFNDTFQLQWMENVHDWVISRQLWWGHQIPAWYNAEG 422

Query: 424 EMYVGEDAPEGDGWTDQEDVLDTWFSALWPFSTMGWPDTEADFKRYFPTSTLVTGYDI 483  
 E+YVGE+APEGD WTQDEVDLDTWFSALWPFSTMGWPD+ DFKRYFPTSTLVTGYDI  
 Sbjct: 423 EIYVGEEAPEGDDWTQDEVDLDTWFSALWPFSTMGWPDTEADFKRYFPTSTLVTGYDI 482

-297-

Query: 484 IFFWVSRMIFQSLEFTGRQPFNSVLIHGLIRDEEGRKMSKSLGNGIDPMDVIEKYGADAL 543  
 IFFWVSRMIFQSLEFTGRQPF NVLIHGLIRDEEGRKMSKSLGNGIDPMDVIEKYGAD+L  
 Sbjct: 483 IFFWVSRMIFQSLEFTGRQPFQNVLIHGLIRDEEGRKMSKSLGNGIDPMDVIEKYGADSL 542

5 Query: 544 RWFLSNGSAPGQDVRFSYEKMDASWNFINKIWNISRYILMNNEGLTLDQARENVEKVVNS 603  
 RWFLSNGSAPGQDVRFSYEKMDASWNFINKIWNISRYILMNNEGLTL+ A NV KV S  
 Sbjct: 543 RWFLSNGSAPGQDVRFSYEKMDASWNFINKIWNISRYILMNNEGLTLEDAESNVAKVAAS 602

10 Query: 604 QVGNVTDRLWILHNLNETVGKVTFENFDKFEFGVAGHILYNFIWEEFANWYVELTKEVLYSD 663  
 + GNVTD+WILHNLNET+ KVTENFDKFEFGVAGHILYNFIWEEFANWYVELTKEVLYSD  
 Sbjct: 603 EAGNVTDQWILHNLNETIAKVTFENFDKFEFGVAGHILYNFIWEEFANWYVELTKEVLYSD 662

15 Query: 664 NEDEKVITRSVLLYTLDQILRLHLPIMPVFVTEEIFGQYAEBSIVLASYPQVNATFENQTA 723  
 NE EKVITRSVLLYTLD+ILRLHLPIMPVFVTEEI+ QYA+GSIV YP V FEN+ A  
 Sbjct: 663 NEAEKVITRSVLLYTLDKILRLHLPIMPVFVTEEIIYAQYAGSIVTVDPVVRPAFENEAA 722

20 Query: 724 HKGVESLKDILIRSVNSRAEVNVAPSKPITILVKTSDSELESFFKDNSNYIKRFTNPETL 783  
 HKGVESLKDILIR+VRN+RAEVNVAPSKPITILVKT+DSELE FF N NYIK FTNPE L  
 Sbjct: 723 HKGVESLKDILIRAVRNARAENVAPSKPITILVKTADSELEDFFNININIKCFTNPEKL 782

25 Query: 784 EISSAIATPELAMSSVITGAEIFLPLADLLNVEEELARLEKELAKWQKELDMVGKKLSNE 843  
 EISSAIA PELAM+S+ITGAEI+LPLADLLNVEEELARL+KELAKWQKELDMVGKKL NE  
 Sbjct: 783 EISSAIAPELAMTSIITGAETYLPLADLLNVEEELARLDKELAKWQKELDMVGKKLGNE 842

Query: 844 RFVANAKPEVVQKEKDKQTDYQTKYDATIARIEEMKKL 881  
 RFVANAKPEVVQKEKDKQ DYQ KYDAT RI EMKK+  
 Sbjct: 843 RFVANAKPEVVQKEKDKQADYQAKYDATQERIAEMKKI 880

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 228

A DNA sequence (GBSx0242) was identified in *S.agalactiae* <SEQ ID 725> which encodes the amino acid sequence <SEQ ID 726>. Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.0669(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 727> which encodes the amino acid sequence <SEQ ID 728>. Analysis of this protein sequence reveals the following:

Possible site: 57  
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----  
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

Identities = 148/191 (77%), Positives = 165/191 (85%)

Query: 14 GEKKKMNIIGQAQSGKMTIGQEIAKQTGMTLFHNHDSIDFVLRFPWPSPDSIALTESI 73

-298-

G + KMN+IIIGAQASGKMTIGQE+A+QTGMTLFHNHDSIDFVLRFPWS +S AL E I  
 Sbjct: 3 GAETKMNLIIIGAQASGKMTIGQEVARQTGMTLFHNHDSIDFVLRFPWSQESTALIERI 62

Query: 74 RFKFFETFAKTGQEMIFTIVIDFNDSDRVVFLLEKIQIVFQSHNQEVLFVELETELSERLK 133  
 RF FFETFAKTGQ+MIFTIVIDFND DV LEKIQ VFQS++QEVLFVEL+T++ ERLK

Sbjct: 63 RFAFFETFAKTGQDMIFTIVIDFNDPNDVAMLEKIQAVFQSYDQEVLFVELKTDIEERLK 122

Query: 134 RNRTENRLKHKPSKRDIKWSESDICSTMDYAI FNPEVAPEALTYHKKINNTCLTATETAY 193  
 RNRTENRLKHKP KR+I+WSE DI STM YA+FNPE P+ LT+Y KINNT LTA ETA

Sbjct: 123 RNRTENRLKHKPLKRNIEWSEQDIQSTMAYAVFNPEEPKTLTHYQKINNTQLTAAETAQ 182

Query: 194 LIIQKINQIKE 204  
 LIIQK+ IKE

Sbjct: 183 LIIQKMTHIKE 193

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 229

A DNA sequence (GBSx0243) was identified in *S.agalactiae* <SEQ ID 729> which encodes the amino acid sequence <SEQ ID 730>. Analysis of this protein sequence reveals the following:

Possible site: 49  
 >>> Seems to have no N-terminal signal sequence

#### ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3614(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04556 GB:AP001510 unknown conserved protein [Bacillus halodurans]  
 Identities = 60/189 (31%), Positives = 102/189 (53%), Gaps = 3/189 (1%)

Query: 7 EIVDNQLPVVETNRLLLRQRKLEDAKEIFEVFKLDEVSYAGFPAPVKSLEEEITYIQEIY 66  
 E + LP +ET RL LR+ +DA I+++ ++V+ + +S+++ ++ +

Sbjct: 4 EDIYGDLPTELETERLRLRKFYKDDAAAIYDYASNEQVTKYVLWETHQSIKDSEAFLA--F 61

Query: 67 PTNLEKEKLPSGYAITLKGD DKVIGSVDFNH-RHEDDIFEIGYLLHPDYWGQGIVPEAAS 125  
 N EK S +AI LK ++++IG+VDF + +D E+GY+L YWGQGI+ EA +

Sbjct: 62 ALNKYDEKDVSPWAIELKRNERMIGTVDFVWVKPKDKTAELGYVLSEPYWGQGIMTEAVN 121

Query: 126 ALVEIGFTLLGLHKEI LGCYDYNKQSQAVARKLGFTLEANIRDRRDAQGKRCGDMRFGLL 185  
 ALVE GF + L +I+ C+ N S V K G E R +G + ++

Sbjct: 122 ALVEFGFNNMELERIQAKCFAENISSARVMEKAGLIYEGTHRRRAIYVKGHRDFKVYAI 181

Query: 186 RSEWEKKRR 194  
 R ++E+K +

Sbjct: 182 REDYEQKHQ 190

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 731> which encodes the amino acid sequence <SEQ ID 732>. Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence

#### ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1864(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

-299-

Identities = 50/58 (86%), Positives = 56/58 (96%)

Query: 137 LHKIELGICYDYNKQSQAVARKLGFTLEANIRDRRDAQGKRCGDMRFGLLRSEWEKKRR 194  
 LHKIELGICYDYNKQSQAVARKLGFTLEAN RDR+D QG+RCGDMRFGLLRSEWE++++  
 5 Sbjct: 1 LHKIELGICYDYNKQSQAVARKLGFTLEANARDRKDVQGRRCGDMRFGLLRSEWEEQKQ 58

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 230**

10 A DNA sequence (GBSx0244) was identified in *S.agalactiae* <SEQ ID 733> which encodes the amino acid sequence <SEQ ID 734>. This protein is predicted to be ribosomal-protein-alanine N-acetyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 54

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4066(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20 A related GBS nucleic acid sequence <SEQ ID 9599> which encodes amino acid sequence <SEQ ID 9600> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

&gt;GP:BAB04418 GB:AP001509 ribosomal-protein-alanine

N-acetyltransferase [Bacillus halodurans]

Identities = 63/185 (34%), Positives = 95/185 (51%), Gaps = 11/185 (5%)

25 Query: 53 KALPKLETDRILRLRQRTVGDVPAMFDYVCLEEVAYPAGLSPIASLEDEYDYFENRYQNL 112  
 K P LET RLILR+ T D ++ Y+ +EV GL P +LED E +Y+++  
 30 Sbjct: 6 KRFPILR+TDRILRLRQRTVGDVPAMFDYVCLEEVAYPAGLSPIASLEDEYDYFENRYQNL 63

Query: 113 EKAKLPSGYGITVKGSDRIIGSCAFN-----HRHEDDVFEICYLLHPDYWGHGYMTEAVA 167  
 + +GIT+KG D +IGSC F+ H + FE+ L YWG G +EA+  
 35 Sbjct: 64 LHEQTGIRWGITLKGQDEIVGSCGFHQWPKHRAEIGFELSCL-----YWGQGIASEAIR 119

Query: 168 ALIEVGFTLLNLHKIEIRCYDYNKQSRRAEKLGFTEATIRDRKDNQDNRCVNLIIYGLL 227  
 A+I+ GF L L +I+ N S+R+ EK GF E +R + +Y LL  
 40 Sbjct: 120 AVIQYGFHLELQRIQALIEPPNIPSRIVEKQGFISEGLLRSYEYTCGKFDDLYMYSLL 179

Query: 228 RSEWE 232

+ +++

Sbjct: 180 KRDFD 184

There is also homology to SEQ ID 732:

Identities = 39/54 (72%), Positives = 44/54 (81%)

45 Query: 179 LHKIEIRCYDYNKQSRRAEKLGFTEATIRDRKDNQDNRCVNLIIYGLLRSEWE 232  
 LHKIE+ CYDYNKQ+ VA KLGFTLEA RDRKD Q RC ++ +GLLRSEWE  
 50 Sbjct: 1 LHKIELGICYDYNKQSQAVARKLGFTLEANARDRKDVQGRRCGDMRFGLLRSEWE 54

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-300-

**Example 231**

A DNA sequence (GBSx0245) was identified in *S.agalactiae* <SEQ ID 735> which encodes the amino acid sequence <SEQ ID 736>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 51
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2719(Affirmative) < succ>
10   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 232**

A DNA sequence (GBSx0246) was identified in *S.agalactiae* <SEQ ID 737> which encodes the amino acid sequence <SEQ ID 738>. Analysis of this protein sequence reveals the following:

```

20   Possible site: 53
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.3250(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
25   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9597> which encodes amino acid sequence <SEQ ID 9598> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 739> which encodes the amino acid sequence <SEQ ID 740>. Analysis of this protein sequence reveals the following:

```

    Possible site: 38
    >>> Seems to have no N-terminal signal sequence

35   ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.3293(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

40 An alignment of the GAS and GBS proteins is shown below:

```

    Identities = 24/55 (43%), Positives = 38/55 (68%)

    Query: 56  LLEGLTANKQDVLKEAGLVSLFAKVSSEADVLALKGIGPAAIKQLVDNGVVFAK 110
              ++ G+ ++  + L   G+ S +AF + +E D+LALKGIGPA +K+LV+NG  F K
45   Sbjct: 77  VVAGIRSDLVETLYAEGIHSAQAFKEWTEKDLLALKGIGPATVKKLVENGASFVK 131

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 233**

A DNA sequence (GBSx0247) was identified in *S.agalactiae* <SEQ ID 741> which encodes the amino acid sequence <SEQ ID 742>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 25
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2901(Affirmative) < succ>
10     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 743> which encodes the amino acid sequence <SEQ ID 744>. Analysis of this protein sequence reveals the following:

```

15     Possible site: 27
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2536(Affirmative) < succ>
20     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

25     Identities = 57/84 (67%), Positives = 73/84 (86%)

Query: 1  MSYEQEFLLKDFEEWLQSQISINQMAMDSAKKVLEEDKDERAADAYIRYESKLDAYRFLQG 60
        MSYE+EFLKDFE+W+++QI +NQ+AM ++++V +ED DERA DA+IRYESKLDAY FL G
Sbjct: 1  MSYEKEFLKDFEDWVKTQIQVNQLAMATSQEVAQEDGDERAKDAFIRYESKLDAYEFLLG 60

30     Query: 61  KFNNYHNQKSFHDLPLDGLFGQRHY 84
        KF+NY N K+FHD+PD LFG RHY
        Sbjct: 61  KFDNYKNGKAFHDIPLDELFGARHY 84

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 234**

A DNA sequence (GBSx0248) was identified in *S.agalactiae* <SEQ ID 745> which encodes the amino acid sequence <SEQ ID 746>. This protein is predicted to be methyltransferase. Analysis of this protein sequence reveals the following:

```

40     Possible site: 61
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2469(Affirmative) < succ>
45     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 747> which encodes the amino acid sequence <SEQ ID 748>. Analysis of this protein sequence reveals the following:

```

50     Possible site: 35
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3352(Affirmative) < succ>

```

-302-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

5       Identities = 26/60 (43%), Positives = 37/60 (61%)  
 Query: 23 LKNERCPHPKLINVLERKLEIILGDQKHILEKDSLISLSPQETHHLRAIENSKFLQIELD 82  
           + E P K+I VLE +L L DQK +L ++SLI++ Q+ HHL A + K LQ+ LD  
 10       Sbjct: 42 ISQETSPRDKVILVLEGQLIFDLEDQKQVLTQESLIAIPAQKVHLEAKTDCKLLQVLLD 101

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 235

15       A DNA sequence (GBSx0249) was identified in *S.agalactiae* <SEQ ID 749> which encodes the amino acid sequence <SEQ ID 750>. This protein is predicted to be integrase (codV). Analysis of this protein sequence reveals the following:

      Possible site: 59  
       >>> Seems to have no N-terminal signal sequence  
 20       ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.3842(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25       No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 236

30       A DNA sequence (GBSx0250) was identified in *S.agalactiae* <SEQ ID 751> which encodes the amino acid sequence <SEQ ID 752>. Analysis of this protein sequence reveals the following:

      Possible site: 22  
       >>> May be a lipoprotein  
 35       ----- Final Results -----  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

40       No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 752 (GBS128) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 23 (lane 5; MW 15kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 32 (lane 4; 2 bands).

45       The GBS128-GST fusion product was purified (Figure 198, lane 2) and used to immunise mice. The resulting antiserum was used for FACS (Figure 288), which confirmed that the protein is immunoaccessible on GBS bacteria.



-303-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 237**

A DNA sequence (GBSx0251) was identified in *S.agalactiae* <SEQ ID 753> which encodes the amino acid sequence <SEQ ID 754>. Analysis of this protein sequence reveals the following:

Possible site: 30  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2940(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 755> which encodes the amino acid sequence <SEQ ID 756>. Analysis of this protein sequence reveals the following:

Possible site: 42  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2518(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 30/90 (33%), Positives = 49/90 (54%), Gaps = 10/90 (11%)

Query: 3 TVAVRVDDQLKDDATELFQSLGLDMSTAVKMFILQSVKTQSIPFEIK-----NKSSV 54  
T+ +RVDD +K A ++ + LG+ MST+ MFL Q + T IPF++ N +

Sbjct: 15 TLNLRVDDSVKSAADDILKRLGIPMSTAIMFLNQIILTGGIPFDVSLPEAPQRVNVDYM 74

Query: 55 SDEEFQNLVETKLKGIKASDPESVNAFF 84  
S E+F ++ T + K +P+ V F+

Sbjct: 75 SQEKFYDKLITSFED--AKTCNPQDVGKIFY 102

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 238**

A DNA sequence (GBSx0252) was identified in *S.agalactiae* <SEQ ID 757> which encodes the amino acid sequence <SEQ ID 758>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

Possible site: 24  
>>> Seems to have no N-terminal signal sequence  
INTEGRAL Likelihood = -2.81 Transmembrane 370 - 386 ( 368 - 388)

----- Final Results -----

bacterial membrane --- Certainty=0.2126(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-304-

A related GBS nucleic acid sequence <SEQ ID 9593> which encodes amino acid sequence <SEQ ID 9594> was also identified. A related GBS nucleic acid sequence <SEQ ID 10773> which encodes amino acid sequence <SEQ ID 10774> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 759> which encodes the amino acid sequence <SEQ ID 760>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -4.57    Transmembrane  354 - 370 ( 353 - 371)

----- Final Results -----
          bacterial membrane --- Certainty=0.2826(Affirmative) < succ>
          bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 344-348

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 64/277 (23%), Positives = 99/277 (35%), Gaps = 31/277 (11%)

Query: 126 SIGNLPDLPGKTTVAFETPVDATATPGDKPAKVVTYPDGSKDVTVDVTVKVVDPRTDADKN 185
      ++ +LP  + TT  E PV          + V          + D+ + T    P  A
Sbjct: 121 AVKDLPASTESTTQPVQETQASASDSMTGDSVTTDSPEETPSSSESPVAPALSE 180

Query: 186 DPAGKDQQVNVGETPKAEDSIGNLPDLPGKTTVAFETPVDATATPGDKPAKVVTYPDGSK 245
      PA  Q    E P  S  P    T  A  ETP + A P  P    +  S+
Sbjct: 181 APA----QPAESEEPSVAASSEETPS--PSTPAAPETPEEPAAPSPSPSEEPSVAAPSE 234

Query: 246 DTVDTVTVKVVDPRTDADKNDPAGKDQQVNVGETPKAEDSIGNLPDLPGKTTVAFETPVD 305
      +T          P  A  + PA ++          T  +          P  P  +  +TP
Sbjct: 235 ETPSPET----PEEPAAPSPQAESEESSVAATTSPS-----PSTPAESET--QTPPAV 281

Query: 306 ATPGDKPAKVVTYPDGSKDVTVDVTVKVVDPRTDADK-----NDPAGKDQQVNGK 355
      DKP+          P  S  + TV+  + +DK          N  +  +  +
Sbjct: 282 TKDSKPSAEEK-PAASSLVSEQTQQPTSKRSSDKKEEQEQSYSPNRSLSRQVRAHES 340

Query: 356 GNKLPAATGENATPFFNVVALTIMSSVGLLSVSKKED 392
      G  LP+TGE A P F +  +T+MS  G L V+K+++
Sbjct: 341 GKYPSTGEKAQPLF-IATMTLMSLFGSLLVTKRQKE 376

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 239

A DNA sequence (GBSx0253) was identified in *S.agalactiae* <SEQ ID 761> which encodes the amino acid sequence <SEQ ID 762>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```

Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.5289(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside  --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 240

A DNA sequence (GBSx0254) was identified in *S.agalactiae* <SEQ ID 763> which encodes the amino acid sequence <SEQ ID 764>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -1.06 Transmembrane 39 - 55 ( 39 - 55)

----- Final Results -----

bacterial membrane --- Certainty=0.1426(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9591> which encodes amino acid sequence <SEQ ID 9592> was also identified.

The protein differs significantly from U58333 in several places:

Query: 157 TKPDGQVDIVNVSLTIYNSSALRDKIDEVKK-----KAED-----PKWDEGSRDK 201  
T PDG D V+V++ + + DK D K KAED P +G+  
Sbjct: 683 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQVNVGETPKAEDSIGNLPDLPGKTTVA 742

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQVNV 258  
+D T D K T D + +VT K++ PR DADKNDPAGKDQQVNV  
Sbjct: 743 FETPVDTA-TPGDKPAKVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQVNV 798

Query: 157 TKPDGQVDIVNVSLTIYNSSALRDKIDEVKK-----KAED-----PKWDEGSRDK 201  
T PDG D V+V++ + + DK D K KAED P +G+  
Sbjct: 841 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQVNVGETPKAEDSIGNLPDLPGKTTVA 900

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQVNV 258  
+D T D K T D + +VT K++ PR DADKNDPAGKDQQVNV  
Sbjct: 901 FETPVDTA-TPGDKPAKVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQVNV 956

Query: 157 TKPDGQVDIVNVSLTIYNSSALRDKIDEVKK-----KAED-----PKWDEGSRDK 201  
T PDG D V+V++ + + DK D K KAED P +G+  
Sbjct: 288 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQVNVGETPKAEDSIGNLPDLPGKTTVA 347

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQVNV 258  
+D T D K T D + +VT K++ PR DADKNDPAGKDQQVNV  
Sbjct: 348 FETPVDTA-TPGDKPAKVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQVNV 403

Query: 157 TKPDGQVDIVNVSLTIYNSSALRDKIDEVKK-----KAED-----PKWDEGSRDK 201  
T PDG D V+V++ + + DK D K KAED P +G+  
Sbjct: 604 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQVNVGETPKAEDSIGNLPDLPGKTTVA 663

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQVNV 258  
+D T D K T D + +VT K++ PR DADKNDPAGKDQQVNV  
Sbjct: 664 FETPVDTA-TPGDKPAKVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQVNV 719

Query: 157 TKPDGQVDIVNVSLTIYNSSALRDKIDEVKK-----KAED-----PKWDEGSRDK 201  
T PDG D V+V++ + + DK D K KAED P +G+  
Sbjct: 446 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQVNVGETPKAEDSIGNLPDLPGKTTVA 505

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQVNV 258  
+D T D K T D + +VT K++ PR DADKNDPAGKDQQVNV  
Sbjct: 506 FETPVDTA-TPGDKPAKVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQVNV 561

Query: 157 TKPDGQVDIVNVSLTIYNSSALRDKIDEVKK-----KAED-----PKWDEGSRDK 201  
T PDG D V+V++ + + DK D K KAED P +G+

-306-

Sbjct: 920 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQVNVGETPKAEDSIGNLPDLPGKTTVA 979

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQVNV 258

+D T D K T D + +VT K++ PR DADKNDPAGKDQQVNV

5 Sbjct: 980 FETPVDTA-TPGDKPAKVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQVNV 1035

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### 10 Example 241

A DNA sequence (GBSx0255) was identified in *S.agalactiae* <SEQ ID 765> which encodes the amino acid sequence <SEQ ID 766>. This protein is predicted to be ara-C-like activator. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.37 Transmembrane 8 - 24 ( 8 - 25)

----- Final Results -----

bacterial membrane --- Certainty=0.1150(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9589> which encodes amino acid sequence <SEQ ID 9590> was also identified.

25 There is homology to SEQ ID 460.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 242

A DNA sequence (GBSx0256) was identified in *S.agalactiae* <SEQ ID 767> which encodes the amino acid sequence <SEQ ID 768>. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1200(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9587> which encodes amino acid sequence <SEQ ID 9588> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 769> which encodes the amino acid sequence <SEQ ID 770>. Analysis of this protein sequence reveals the following:

Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0679(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-307-

An alignment of the GAS and GBS proteins is shown below:

Identities = 135/176 (76%), Positives = 161/176 (90%)

```

5  Query: 1  MSYMVKDRQIQTKVAIYNFISLLQENDYSKITVQDVIGLANVGRSTFYSHYESKEVLL 60
    +S M KDRQI+KTK AIY+AFI+LLQ+ +YSKITV+D+I LANVGRSTFY+HYESKE+LL
    Sbjct: 1  VSDMTKDRQIKKTKTAIYSAFIALLQKKEYSKITVRDMITLANVGRSTFYAHYESKEMLL 60

10 Query: 61  KELCEDLFHHLFKQGRDVTFFEEYLVHILKHFEQNQDSIATLLLSDDPYFLLRFRSELEHD 120
    KELCE+LFHHLF+Q R+VTFE+YLVHILKHFEQN+DSIATLLLS+DPYFLLRF++ELEHD
    Sbjct: 61  KELCEEELFHHLFRQKRNVTFEDYLVHILKHFEQNKDSIATLLLSNDPYFLLRFKNELEHD 120

    Query: 121 VYPRLREYITKVDIPEDFLKQFLLSSFIETLKWHLHQRQKMTVEDLLKYLYTMVE 176
    VYP LR +YI K IPE FLKQF+LSSFIETLKWHLHQRQ+M+ +LLKYLYL +++
15  Sbjct: 121 VYPNLRCKYIDKTTIPEVFLKQFVLLSSFIETLKWHLHQRQMSANELLYLELIK 176

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 243

20 A DNA sequence (GBSx0257) was identified in *S.agalactiae* <SEQ ID 771> which encodes the amino acid sequence <SEQ ID 772>. Analysis of this protein sequence reveals the following:

Possible site: 29  
>>> Seems to have no N-terminal signal sequence

```

25  ----- Final Results -----
        bacterial cytoplasm --- Certainty=0.3573(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 244

35 A DNA sequence (GBSx0258) was identified in *S.agalactiae* <SEQ ID 773> which encodes the amino acid sequence <SEQ ID 774>. Analysis of this protein sequence reveals the following:

Possible site: 35  
>>> Seems to have a cleavable N-term signal seq.

```

40  INTEGRAL    Likelihood = -10.19    Transmembrane  112 - 128 ( 107 - 131)
    INTEGRAL    Likelihood = -8.07     Transmembrane   77 - 93 ( 71 - 97)
    INTEGRAL    Likelihood = -6.10     Transmembrane  144 - 160 ( 138 - 165)
    INTEGRAL    Likelihood = -3.03     Transmembrane  165 - 181 ( 164 - 182)

```

```

45  ----- Final Results -----
        bacterial membrane --- Certainty=0.5076(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 775> which encodes the amino acid sequence <SEQ ID 776>. Analysis of this protein sequence reveals the following:

Possible site: 35  
>>> Seems to have an uncleavable N-term signal seq

```

    INTEGRAL    Likelihood = -9.13     Transmembrane  112 - 128 ( 107 - 130)

```

-308-

```

INTEGRAL    Likelihood = -5.89    Transmembrane  144 - 160 ( 138 - 163)
INTEGRAL    Likelihood = -5.47    Transmembrane   7 - 23 (   6 - 29)
INTEGRAL    Likelihood = -3.50    Transmembrane  77 - 93 (  74 - 94)
INTEGRAL    Likelihood = -2.07    Transmembrane 166 - 182 ( 165 - 183)

```

```

----- Final Results -----

```

```

      bacterial membrane --- Certainty=0.4652(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

Identities = 212/287 (73%), Positives = 245/287 (84%)

```

Query: 1   MTSNKKVAIAFILNISFSVLEFIFGSLFFSGAILADAVHDFGDAIAIGISATLEKKS KKD 60
          M ++KKV I FILN+SFS++EFIFG+LFFSGAILADAVHDFGDAIAIGISA LE+K+ K
Sbjct: 1   MPASKKVTIIFILNLSFSLIEFIFGTLFFSGAILADAVHDFGDAIAIGISAILERKAVKK 60

Query: 61  EDTIFSLGYKRFSLLGALITSLILISGSILVMENIPKLWHPTPVNYHGMFILAVIAIII 120
          E FSLGYKRFSLLGAL T+LILISGS+LVMIE IPKLWHPT VNY GMF+LA+ AIII
Sbjct: 61  ESPNFSLGYKRFSLLGALTNNLILISGSLLVMIE TIPKLWHPTIVNYDGMFVLAIFAIII 120

Query: 121 NGLASFILHSGQSKHEILSLHFLEDILGWLAIIVISLILNWKPLYILDPLLSVAISTFI 180
          NG ASFI+HS Q+K+EEILSLHFLEDILGWLAI++SLIL WKP YILDPLLS+AI++FI
Sbjct: 121 NGFASFI+HSNQTKNEEILSLHFLEDILGWLAI+ILSLILKWKWPYILDPLLSIAIASFI 180

Query: 181 LSKALPKLLSTLKLFLDGVPSIDYAAALHDELKGLSQVRSTNQLNIWSMDGIDNRRAIHC 240
          LSKALPKL++T +FLDGVPSIDY LH EL L + S+NQLN+WSDMGID+RA IHC
Sbjct: 181 LSKALPKLVATANIFLDGVPSIDYCTLHHELSQLPHIVSVNQLNVWSMDGIDHRATIHC 240

Query: 241 CLNQLISEKDCKRAIRITICQHYKINDVTVEIDYSLREHQNHCKPLKN 287
          CL + +EK CK++IR ICQ Y IN VTVEID SL EHQ+HC L +
Sbjct: 241 CLRESTTEKHCKKSIRLICQRYNINSVTVEIDTSLNEHQHHCSSLSS 287

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 245

A DNA sequence (GBSx0259) was identified in *S.agalactiae* <SEQ ID 777> which encodes the amino acid sequence <SEQ ID 778>. Analysis of this protein sequence reveals the following:

```

Possible site: 48
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.22    Transmembrane  221 - 237 ( 221 - 237)

```

```

----- Final Results -----

```

```

      bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

There is also homology to SEQ ID 780.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 246

A DNA sequence (GBSx0260) was identified in *S.agalactiae* <SEQ ID 781> which encodes the amino acid sequence <SEQ ID 782>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -2.50    Transmembrane   2 - 18 (   1 - 18)

```

-309-

## ----- Final Results -----

5           bacterial membrane --- Certainty=0.1999(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

10   Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 247**

A DNA sequence (GBSx0261) was identified in *S.agalactiae* <SEQ ID 783> which encodes the amino acid sequence <SEQ ID 784>. This protein is predicted to be dehydrogenase (Zn-dependent). Analysis of this protein sequence reveals the following:

15   Possible site: 15  
       >>> Seems to have no N-terminal signal sequence  
           INTEGRAL   Likelihood = -3.77   Transmembrane   171 - 187 ( 170 - 187)

20   ----- Final Results -----  
           bacterial membrane --- Certainty=0.2508(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

25   >GP:AAG20655 GB:AE005134 alcohol dehydrogenase; Adh2 [Halobacterium  
       sp. NRC-1]  
       Identities = 169/348 (48%), Positives = 232/348 (66%), Gaps = 9/348 (2%)

30   Query: 1   MKVATFIEPGKMVITDTPKPVIEQETDAVIKIVRACVCGSDLWWYRGISKRESGSFAGHE 60  
           M+ A + PG++ + + PKP IE   DAVI++   VCGSDLW+YRG S RE+GS GHE  
   Sbjct: 1   MRAAVYQGPGEIAVEEVPKPDIESPEDAVIRVTHTAVCGSDLWYRGGSDREAGSRVGHE 60

35   Query: 61   AIGIVEEVGTVKVTDVSKGDFVIVPFTHGCGQCPSCKAGFDGNCNTNHQA---AKNVGYQGO 117  
           +GIVEEVG VT V+ GD VI PF CG+C C+ G +C ++   N G QG+  
   Sbjct: 61   PMGIVEEVGDDVTSVAPGRVIAFPFAISCGECEFCRQGLYTSCVEDESWGSEANGGGQGE 120

40   Query: 118 YLRYTNANWALVKIPGQPSDYDNETLNSLLTSDVMATGYHAAATAEVKEGDTVVMGDG 177  
           Y++ A+ LV++P + +D D + L SLL L+DVM TG+HAA +A V EGDV VV+GDG  
   Sbjct: 121 YVKCFPADGTLVRVPDRYAD-DEDVLESLLPLTDVMGTGHHAASAGVGEGDTAVVVGDG 179

45   Query: 178 AVGLCGVIAAKMLGANRIIAMS RHKDRQELALTFGATDIVEERGDEAVKRVLDLTNQAGA 237  
           AVGLCGV+AA+ LGA RIIAM H+DR ELA FGATD + RGD+A++R DLT+ GA  
   Sbjct: 180 AVGLCGVLAAQRLGAERI IAMGHEDRLELAAEFGATDTISARGDDAIERARDLTH-GGA 238

50   Query: 238 DAVLECVGTEQSVDTATQIARPGAVIGRVGIP---QNPDMNTNNLFWKNIGLRGGIASVT 294  
           + V+ECVG ++D+A IARPG +G VG+P ++ ++ +F NI +RGG+A V  
   Sbjct: 239 NHVMECVGAASAMDSAIAIARPGGTVGYPYGVGVEDGGLDVFTMFSDNITIRGGVAPVR 298

55   Query: 295 TFDKSVLLDAVLTHKINPGLVFTKSFVLDDIQKAYEAMDKRDAIKSLV 342  
           + + ++ D VL ++P +FTK+ LD + + Y AMD R+AIK LV  
   Sbjct: 299 AYAEELMAD-VLQGTLDPSPIFTKTVDLDGVPEGYAAMDREAIKVLV 345

There is also homology to SEQ ID 786.

55   A related sequence was also identified in GAS <SEQ ID 9145> which encodes the amino acid sequence <SEQ ID 9146>. Analysis of this protein sequence reveals the following:

Possible site: 23

-310-

>>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -5.41 Transmembrane 170 - 186

5 ----- Final Results -----  
           bacterial membrane --- Certainty=0.3166(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 An alignment of the GAS and GBS proteins is shown below:

Identities = 121/353 (34%), Positives = 182/353 (51%), Gaps = 16/353 (4%)

Query: 1 MKVATFIEPGKMVITDTPKPVIEQETDAVIKIVRACVCGSDLWYRG-ISKRESGSFAGH 59  
 MK AT++ G + + D PKPVI + TDA++++V+ +CG+DL G + + G+ GH  
 15 Sbjct: 15 MKAATYLSGTGNLQIDKPKPVIKPTDAIVQLVKTTCGTLHLGGDVPACKEGTILGH 74

Query: 60 EAIGIVEEVGTVKVDVSKGDFVIVPFTTHGCGQCPCKAGFDGNC'TNHQAANK---VG YQG 116  
 E IGIV+EVG VT+ GD VI+ C C CK G +C + G Q  
 20 Sbjct: 75 EGIGIVKEVGDAVTNFKIGDKVIISCVTSCHTCYCKRGLSSHQDGGWILGHLINGTQA 134

Query: 117 QYLRYTNANWALVKIPGQPSDYDNETLNSLLTSLSDVMATGYH-AAATAEVKEGDTVVMVG 175  
 +Y+ +A+ +L P D +L+ LSD++ T Y + VK GD V ++G  
 25 Sbjct: 135 EYVHIPHADGSLYHAPDTIDD-----EALVMLS DILPTS YEIGVLP SHVKPGDNVCIVG 188

Query: 176 DGAVGLCGVIAAKMLGANRIIAMS SRHKDRQELALTFGATDIVEERGDEAVKRVL-DLTNQ 234  
 G VGL ++ + II + ++R E A TFGAT + E VK ++ D+TN  
 30 Sbjct: 189 AGPVGLAALLTVQFFSPANIIMVDLSQNRLEAAKTFGATHTICSGSSEEVKAIIDITNG 248

Query: 235 AGADAVLECVGTEQSVDTATQIARPGAVIGRVGIPQNP-DMNTNNLFWKNIGLRGGIASV 293  
 G D +ECVG + D +I G I VG+ P D N + L+ KNI L G+ +  
 35 Sbjct: 249 RGVDISMECVGYPATFDICQKIISVGGHIANVGVHGPVDFNLDELWIKNITLNTGLVNA 308

Query: 294 TTFDKSVLLDAVLTHKINPGLVFTKSFVLDDIQAYEAMDKRDAIKSL-VIVD 345  
 T + +LL+ + T KI+ + T F L +++KAYE A +L VI+D  
 35 Sbjct: 309 NTTE--MLLNVLKTGKIDATRLITHHFKLSEVEKAYETFKHAGANNALKVIID 359

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 248

40 A DNA sequence (GBSx0262) was identified in *S.agalactiae* <SEQ ID 787> which encodes the amino acid sequence <SEQ ID 788>. Analysis of this protein sequence reveals the following:

Possible site: 31  
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.2169(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD36075 GB:AE001762 hypothetical protein [Thermotoga maritima]  
 Identities = 55/128 (42%), Positives = 72/128 (55%), Gaps = 8/128 (6%)

Query: 8 IFPKGEKNPYGEFFIGQSYLAALAKSPDG--NVSVGNVTFEAGCRNNWHVHLDGYQILLV 65  
 IF +G K +FF G ++ L +G N V +V FE G R +WH H G QIL+V  
 55 Sbjct: 5 IFERGSKGS-SDFFTGNVWVKMLVTDENGVFNTQVYDVVFEPGARTHWHSHPGG-QILIV 62

Query: 66 TEGSGWYQEEGKEAVSLKPGDVIVITDKGVRHWHGAKKDSEFAHIAITA---GKSEFYEA 121  
 T G G+YQE GK A LK GDV+ V HWHGA D E HI I+ G +E+ +  
 60 Sbjct: 63 TRGKGFYQERGKPARILKKGDVVEIPPVNVHWHGAAPDEELVHIGISTQVHLGPAEWLGS 122



-311-

Query: 122 VSDEEYSR 129  
 V++EEY +  
 Sbjct: 123 VTDEEYRK 130

5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 249

A DNA sequence (GBSx0263) was identified in *S.agalactiae* <SEQ ID 789> which encodes the amino acid  
 10 sequence <SEQ ID 790>. This protein is predicted to be gamma-carboxymuconolactone decarboxylase.  
 Analysis of this protein sequence reveals the following:

Possible site: 49  
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.4089(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA20070 GB:AL031155 3-oxoadipate enol-lactone  
                   hydrolase/4-carboxymuconolactone decarboxylase  
                   [Streptomyces coelicolor A3(2)]  
 Identities = 33/93 (35%), Positives = 59/93 (62%), Gaps = 1/93 (1%)

25 Query: 11 QLEEFAPFARYNDILFGEVWAKEDHLTDKTRSIITISALISGGNLEQLEHHLQFAKQN 70  
           Q +EF+ +F +           +GE+W +   L ++RS +T++AL++GG+L++L HL+ A +N  
 Sbjct: 349 QADEFSGDFQEFTRYANGEIWDRPG-IDRRSRSCVTLTALVAGGHLDELAPHLRAALRN 407

30 Query: 71 GVTKEEADIITHLAFYVGWPKAWSAFNKAKEI 103  
           G+T EI +++   A Y G P A AF A+++  
 Sbjct: 408 GLTPGEIKEVLLQAAVYCGVPAANGAFRVAQQV 440

No corresponding DNA sequence was identified in *S.pyogenes*.

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 250

A DNA sequence (GBSx0265) was identified in *S.agalactiae* <SEQ ID 791> which encodes the amino acid  
 sequence <SEQ ID 792>. Analysis of this protein sequence reveals the following:

40 Possible site: 44  
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.5529(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 251

A DNA sequence (GBSx0266) was identified in *S.agalactiae* <SEQ ID 793> which encodes the amino acid sequence <SEQ ID 794>. This protein is predicted to be probable transcriptional regulator. Analysis of this protein sequence reveals the following:

Possible site: 58  
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9585> which encodes amino acid sequence <SEQ ID 9586> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG08263 GB:AE004901 probable transcriptional regulator  
[Pseudomonas aeruginosa]  
Identities = 36/148 (24%), Positives = 68/148 (45%), Gaps = 22/148 (14%)  
  
Query: 5 QIVEKPAMILAG-----VTLENVKSNQEGIQQAIGICKTQPDFRFD 45  
+IVE+PA + G + E+ + + + GIC QP+ F  
Sbjct: 123 RIVERPAFSVVGMEYFGSAPGDTIGQLWERFIPREHEIAGKHDPEVSYGICAQQPNGEFH 182  
  
Query: 46 YSATYQVETSVQAPKGLEIIRIPSATYAVISVKGPMPSLQETWRKIIQGGFFQENNLKPA 105  
Y A ++V+ P+G+ ++P+ YAV + KG P + E+++ I E L+P  
Sbjct: 183 YVAGFEVQEGWPFVPEGMVRFQVPAQKYAVFTHKGTAP-QIAESFQAIYSHLLAERGLEPK 241  
  
Query: 106 NSPNLEIYSSQH--PQD TDYQMEIWLAI 131  
+ E Y + P D + Q++++ I  
Sbjct: 242 AGVDFEYYDQRFGRGLDPNSQVDLYIPI 269

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 252

A DNA sequence (GBSx0267) was identified in *S.agalactiae* <SEQ ID 795> which encodes the amino acid sequence <SEQ ID 796>. Analysis of this protein sequence reveals the following:

Possible site: 24  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.0887(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB84919 GB:AE000825 conserved protein [Methanothermobacter  
thermoautotrophicus]  
Identities = 42/130 (32%), Positives = 71/130 (54%), Gaps = 3/130 (2%)

-313-

Query: 1 MITQEMKEIINSQLAMVATVDAKGQPNIGPKRSMRLWDDKTFIYNENTDGQTRINIEDNG 60  
 M+T EM + I +L VAT D +G PN+ P R D++T + +N +T N+ +N  
 Sbjct: 1 MMTPEMMDAIEKELVVFVATADEEGTPNVVPIGFARPLDERTILIADNYMKKTIRNLHENP 60

5 Query: 61 KIEIAFVDRERLLGYRFVGTAEIQTEGTYEEAAKKWAEGRMG--VPKAVGIIHVERIFNL 118  
 +I + R Y+F GT EI G Y++ +WA+ M PK+ ++ VE I+++  
 Sbjct: 61 RIAL-IPQNARECPYQFKGTVEIFKSGKYFDMVVEWAQNVMTLEPKSAILMTVEEITYSV 119

10 Query: 119 QSGANAGKEI 128  
 + G AG+++  
 Sbjct: 120 KPGPEAGEKV 129

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 797> which encodes the amino acid sequence <SEQ ID 798>. Analysis of this protein sequence reveals the following:

15 Possible site: 24  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.0789(Affirmative) < succ>  
 20 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 123/128 (96%), Positives = 127/128 (99%)

25 Query: 1 MITQEMKEIINSQLAMVATVDAKGQPNIGPKRSMRLWDDKTFIYNENTDGQTRINIEDNG 60  
 MITQEMK++IN+QLAMVATVDAKGQPNIGPKRSMRLWDDKTFIYNENTDGQTRINIEDNG  
 Sbjct: 1 MITQEMKDLINNQLAMVATVDAKGQPNIGPKRSMRLWDDKTFIYNENTDGQTRINIEDNG 60

30 Query: 61 KIEIAFVDRERLLGYRFVGTAEIQTEGTYEEAAKKWAEGRMGVPAVGIIHVERIFNLQS 120  
 KIEIAFVDRERLLGYRFVGTAEIQTEG TYEEAAKKWA+GRMGVPAVGIIHVERIFNLQS  
 Sbjct: 61 KIEIAFVDRERLLGYRFVGTAEIQTEGAYEEAAKKWAQGRMGVPAVGIIHVERIFNLQS 120

35 Query: 121 GANAGKEI 128  
 GANAGKEI  
 Sbjct: 121 GANAGKEI 128

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### 40 Example 253

A DNA sequence (GBSx0268) was identified in *S.agalactiae* <SEQ ID 799> which encodes the amino acid sequence <SEQ ID 800>. Analysis of this protein sequence reveals the following:

Possible site: 22  
 >>> Seems to have a cleavable N-term signal seq.

45 INTEGRAL Likelihood = -5.47 Transmembrane 1028 -1044 (1027 -1048)

----- Final Results -----  
 bacterial membrane --- Certainty=0.3187(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

!GB:AF054892 surface antigen BspA [Bacteroides forsy...  
 !GB:AF054892 surface antigen BspA [Bacteroides forsy...  
 55 !GB:AF054892 surface antigen BspA [Bacteroides forsy...  
 !GB:AF054892 surface antigen BspA [Bacteroides forsy...  
 !GB:AF054892 surface antigen BspA [Bacteroides forsy...  
 >GP:AAC82625 GB:AF054892 surface antigen BspA [Bacteroides

forsythus]

Identities = 143/566 (25%), Positives = 243/566 (42%), Gaps = 52/566 (9%)

5 Query: 95 VPKAKPEVTQEASNSSNDASKVEVPKQDTASKKETLETSTWEAKDFVTRGDTLVG----F 150  
 +P + + + A + + +P TA + L T + + T +G F  
 Sbjct: 120 IPNSVTTIGEWAFKGCGLKSLTLPNSLTAIGQSALSGCTGLTSITIPNSVTTIGEWAF 179

10 Query: 151 SKSGINKLSQTSHLVLP SHAA--DGTQLTQVASFAPDPKKTAAEYTSRLGENGKPSRL 208  
 SG+ ++ + L +A LT + PD T I E + G +G S  
 Sbjct: 180 GCSGLTSITFPNSLTAIGESAFYGCGLTSIT----LPDALTTIGESAFK-GCSGLKSIT 234

15 Query: 209 DIDQKEIIDEGEIFNAYQLTKLTIPNGYKSIGQDAFVDNKNIAEVNLPESLETISDYAFA 268  
 + I E ++ LT +T+P+ +IG+ AF + + P SL TI + AF  
 Sbjct: 235 FPNSLTTIGESAFYDCGLTSITLPDALTTIGRSAFYGCGLKSITFPNSLTTIGESAFY 294

20 Query: 269 HM-SLKQVKLPDNLKVIGELAFFDNQIGGKLYLPRHLIKLAERAFKSNRIQTVEFLGSKL 327  
 + SL + +P+++ IG AF+ + LP L + ERAF + + T + + +  
 Sbjct: 295 NCGSLTSITIPNSVTTIGRSAFYGCGLKSITLPDGLTTIEERAFYNCVLTSTIPNSV 354

25 Query: 328 KVIGESAFQD--NNLRNMLPDGLEKIESEAFNGPGEHYNNQVLRTRTGQNPQLATE 386  
 IGE++F + L+++ LPDGL IE AF N L + T N E  
 Sbjct: 355 ATIGESAFYGCGLKSITLPDGLTTIEWGAFY-----NCGALTSITIPNSVSTIGE 405

30 Query: 387 NTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNSVTGFS---NKGLQKVRNKNLEIPKQH 443  
 + + +L T D ++ D +++ ++G G + V K ++ K+  
 Sbjct: 406 SAFYGCGLKALDVTVAWDTPIDIQRD-VFRELTLSGIRLHVPAGKKTVEAK--DVWKEF 461

35 Query: 444 NGITITEIGDNAFRNVDFQSKTLRKYDLEEIKLPSTIRKIGAFQSNLKSFEASEDLE 503  
 N + + G + N D +KTL + P T + + FA ++ L  
 Sbjct: 462 NIVEDDDFGGLQW-NYDAATKLTITN----PTPDTPKPMPNFATPDQLW----- 507

40 Query: 504 EIKEGAFMNNRIGTLDLKDLIKIGDAAFH-INHIYAIVLPESVQEIGRSAFRQNGALHL 562  
 GAF I + + D + +GD AF + + +I LP+SV IG+SAF L  
 Sbjct: 508 ----GAFQKE-IQKITIGDGVTSVGDFAFSGCDALKSITLPKSVTTIGQSAFSGCWDLRS 562

45 Query: 563 MFIGNKVKTIGEMAFLSNKLESVNLSQKQLKTIEVQAFS-DNALSEVVLPPNLQTIREE 621  
 + + + V TIGE AF + LE +++ K + I + F +L+ + LP L I ++  
 Sbjct: 563 LTLPDGVNTIGEKAFY-DCLELTSITIPKSVTAIGQETPHYCVSLTSLTLPDALTAIGKK 621

50 Query: 622 AF-KRNHLKEVKGSSTLSQITFNAFD 646  
 AF N L V +++ I NAFD  
 Sbjct: 622 AFYSCNALTSVTFPKSITTIGENAFD 647  
 Identities = 109/407 (26%), Positives = 175/407 (42%), Gaps = 48/407 (11%)

55 Query: 222 FNAYQLTKLTIPNGYKSIGQDAFVDNKNIAEVNLPESLETISDYAFAHMS-LKQVKLPDN 280  
 F+ LT +T+PN +IG AF + + +P S+ TI ++AF S LK + LP++  
 Sbjct: 87 FSDCALTSVTLPNLSLTAIGDHAFKGCGLTSITIPNSVTTIGEWAFKGCGLKSITLPNS 146

60 Query: 281 LKVIGELAFFDNQIGGKLYLPRHLIKLAERAFKSNRIQTVEFLGSKLKVIGESAFQD--NN 339  
 L IG+ A + +P + + E AF T + L IGE++F  
 Sbjct: 147 LTAIGQSALSGCTGLTSITIPNSVTTIGEWAFGCSGLTSITFPNSLTAIGESAFYGC GA 206

65 Query: 340 LRNVMLPDGLEKIESEAFNGPGEHYNNQVLRTRTGQNPQLATENTYVNPDKSLWRA 399  
 L ++ LPD L I AF G G L++ T N E+ + +  
 Sbjct: 207 LTSITLPDALTTIGESAFKGCGLSITFPNSLTTIGESAFYDCGLTSIT 257

Query: 400 TPDMDYTKWLEEDFTYQKNSVTGFSNKGKQKVRNKNLEIPKQHNIGITITEIGDNAFRNV 459  
 PD ++T K++ P ++T IG++AF N  
 Sbjct: 258 LPD-----ALTTIGRSAFYGCGLKSTTFPN-----SLTTIGESAFYNC 296

Query: 460 DFQSKTLRKYDLEEIKLPSTIRKIGAFQSNLKSFEASEDLEEIKEGAFMNNRIGT- 517  
 L I +P+++ IG AF + LKS + L I+E AF N + T  
 Sbjct: 297 G-----SLTSITIPNSVTTIGRSAFYGCGLKSITLPDGLTTIEERAFYNCVLT 347

Query: 518 LDLKDKLIKIGDAAFH-INHIYAIVLPESVQEIGRSAFRQNGALHLMFIGNKVKTIGEMA 576  
 + + + + IG++AF+ + + +I LP+ + I AF GAL + I N V TIGE A  
 Sbjct: 348 ITIPNSVATIGESAFYGCGLKSITLPDGLTTIEWGAFYNCALTSITIPNSVSTIGESA 407

-315-

- Query: 577 FLS-NKLESVNLSEQQLKTIEVQAFSDNALSEVVL--PPNLQTIRE 620  
F L+ V ++ + I+ F + LS + L P +T+ E  
Sbjct: 408 FYGCGALKDVTVAWDTPI-DIQRDVFRELTLGIRLHVPAGKKTVE 453  
Identities = 111/465 (23%), Positives = 185/465 (38%), Gaps = 56/465 (12%)
- 5 Query: 141 VTRGDTLVGFSKSGINKLSQTSHLVLP SHAADGTQLTQVASFAF-----TPDKKT 190  
+T D L +S S + P+ LT + AF PD T  
Sbjct: 210 ITLPDALTTIGESAFKGCGLKSITFPN-----SLTTIGESAFYDCGALTSITLPDAL 263
- 10 Query: 191 AIAEYTSRLGKPSRLDIDQKEIIDEGEIFNAYQLTKLTIPNGYKSIGQDAFVDNKNI 250  
I ++ G+G S + I E +N LT +TIPN +IG+ AF +  
Sbjct: 264 TIGR-SAFYGCGLKSITFPNSLTIGESAFYNGSLTSITIPNSVTIGRSAFEYGCGL 322
- 15 Query: 251 AEVNLPESETISDYAFAHMS-LKQVKLPDNLKVIGELAFFDNQIGGKLYLPRHLIKLAE 309  
+ LP+ L TI + AF + L + +P+++ IGE AF+ + LP L +  
Sbjct: 323 KSITLPDGLTTIEERAFYNGVLTISITIPNSVATIGESAFYGCGLKSITLPDGLTTIEW 382
- 20 Query: 310 RAFKSNRIQTVEFLGSKLVIGESAFQD-NNLRNVMLP-DGLEKIESEAF-----TGNPG 362  
AF + T + + + IGE++F L++V + D I+ + F +G  
Sbjct: 383 GAFYNGGALTSITIPNSVTIGESAFYGCGLKDVTVAWDTPIDIQRDVFRELTLGIRL 442
- 25 Query: 363 DEHYNNQVVLRTTRTGQNPQLATEN-----TYVNPDKSLWRATPDMYTKWLEEDFTY 415  
+ V + + ++ Y K+L P D K + +F  
Sbjct: 443 HVPAGKKTVEYAKDVWKEFNIVEDDDDFGLQWNYDAATKTLTITNPTDTPKPM-PNFAT 501
- 30 Query: 416 QKNSVTGFSNKGKLVRRNKNLEIPKQHNIGITITEIGNAFRNVDFQSKTLRKYDLEIK 475  
+ + G K +QK+ G +T +GD AF D L+ I  
Sbjct: 502 PNDQLWGAFQKEIQKIT-----IGDGVTSVGDFAFSGCD-----ALKSIT 541
- 35 Query: 476 LPSTIRKIGAFAFQSN-NLKSFEASEDLEEIKEGAFMN-NRIGTLDLKD KLIKIGDAAFH 533  
LP ++ IG AF +L+S + + I E AF + + ++ + + IG FH  
Sbjct: 542 LPKSVTTIGQSAFSGCWDLRSLTLPDGVNTIGEKAFYDCLELTSITIPKSVTAIGQETFH 601
- 40 Query: 534 -INHIYAIVLPESVQEIGRSAFRQNGALHLMFIGNKVKTIGEMAF 577  
+ ++ LP+++ IG+ AF AL + + TIGE AF  
Sbjct: 602 YCVSLTSLTLPDALTAIGKKAFFYSCNALTSVTFPKSITITIGENAF 646  
Identities = 98/351 (27%), Positives = 152/351 (42%), Gaps = 53/351 (15%)
- 45 Query: 315 NRIQTVEFLGSKLVIGESAFQDNNLRNVMLPDGLEKIESEAF TGNPGDEHYNNQVVLRT 374  
++IQTV +G + +G +F D L +V LP+ L I AF G G L +  
Sbjct: 68 SKIQTVT-IGDGVTSVGNNAFSDCALTSVTLPNLSLTAIGDHAFKGCGLS-----LTS 117
- 50 Query: 375 RTGQNPQLATENTYVNPDKSLWRATPDMYTKWLEEDFTYQKNSVTGFSNKGKLVRRN 434  
T P+ + T + S ++ NS+T L  
Sbjct: 118 IT--IPNSVTTIGEWAFKGCGLKSIT-----LPNSLTAIGQSALSGCTGL 161
- 55 Query: 435 KNLEIPKQHNIGITITEIGNAF-----RNVDFQSKTLRKYD-----LEEIKLPSTI 480  
++ IP ++T IG+ AF ++ F + + L I LP +  
Sbjct: 162 TSITIPN-----SVTTIGEWAFGCGSLTSITFPNSLTAIGESAFYGCGLTSITLPDAL 216
- 60 Query: 481 RKIGAFAFQS-NNLKSFEASEDLEEIKEGAFMN-NRIGTLDLKD KLIKIGDAAFH-INHI 537  
IG AF+ + LKS L I E AF + + ++ L D L IG +AF+ + +  
Sbjct: 217 TTIGESAFKGCGLKSITFPNSLTIGESAFYDCGALTSITLPDALTTIGRSAFEYGCGL 276
- 65 Query: 538 YAIVLPESVQEIGRSAFRQNGALHLMFIGNKVKTIGEMAFSL-NKLESVNLSEQQLKTI 596  
+I P S+ IG SAF G+L + I N V TIG AF + L+S+ L + L TI  
Sbjct: 277 KSITFPNSLTIGESAFYNGSLTSITIPNSVTIGRSAFEYGCGLKSITLPD--GLTTI 334
- 60 Query: 597 EVQAFSD-NALSEVVLPPNLQTIREEAFKR-NHLKEVKGSSSTLSQITFNAF 645  
E +AF + L+ + +P ++ TI E AF + LK + L+ I + AF  
Sbjct: 335 EERAFYNGVLTISITIPNSVATIGESAFYGCGLKSITLPDGLTTIEWGAF 385  
Identities = 78/282 (27%), Positives = 123/282 (42%), Gaps = 46/282 (16%)
- 65 Query: 111 NDASKVEVPKQDTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVLP-- 168  
N+AS E+P SK +T VT GD + + + TS + LP+  
Sbjct: 56 NNAS--EIPWHSLSKSIQT-----VTIGDGVTSVGNNAFSDCALTS-VTLPNLSL 101
- Query: 169 -----HAADG-----TQLTQVASFAFT-----PDKKTAIAEYTSRLGENG 203

-316-

HA G +T + +AF P+ TAI + ++ G G  
 Sbjct: 102 TAIGDHAFKGCGLTSITIPNSVTTIGEWAFKGCGLKSITLPSNLTAIGQ-SALSGCTG 160

Query: 204 KPSRLDIDQKEIIDEGEIFNAYQLTKLTIPNGYKSIGQDAFVDNKNIAEVLNPESLETIS 263  
 S + I E F LT +T PN +IG+ AF + + LP++L TI  
 Sbjct: 161 LTSITIPNSVTTIGEWAFKGCGLTSITIPNSLTAGESAIFYGCGALTSITLPDALTTIG 220

Query: 264 DYAFAHMS-LKQVKLPDNLKVIGELAFFDNQIGGKLYLPRHLIKLAERAFKS-NRIQTVE 321  
 + AF S LK + P++L IGE AF+D + LP L + AF + +++  
 Sbjct: 221 ESAFKGCGLKSITIPNSLTIGESAIFYDCGALTSITLPDALTTIGRSIFYGCGSLKSIT 280

Query: 322 FLGSKLVIGEEASFQD-NNLRNVMLPDGLEKIESEAFNG 362  
 F S L IGE++F + +L ++ +P+ + I AF G G  
 Sbjct: 281 FPNS-LTTIGESAIFYNCGSLTSITIPNSVTTIGRSIFYGCSG 321  
 Identities = 43/144 (29%), Positives = 70/144 (47%), Gaps = 4/144 (2%)

Query: 220 EIFNAYQ--LTKLTIPNGYKSIGQDAFVDNKNIAEVLNPESLETISDYAFAHM-SLKQVK 276  
 +++ A+Q + K+TI +G S+G AF + + LP+S+ TI AF+ L+ +  
 Sbjct: 505 QLWGAFQKEIQKITIGDGVTSVGDFAFSGCDALKSITLPKSVTTIGQSAFSGCWDLRSLT 564

Query: 277 LPDNLKVIGELAFFDNQIGGKLYLPRHLIKLAERAFKSNRIQTVEFLGSKLVIGEEASFQ 336  
 LPD + IGE AF+D + +P+ + + + F T L L IG+ +F  
 Sbjct: 565 LPDGVNTIGEKAFYDCLELTSITIPKSVTAIGQETFHVCVSLTSLTLPDALTAIGKKAFY 624

Query: 337 D-NNLRNVMLPDGLEKIESEAFNG 359  
 N L +V P + I AF G  
 Sbjct: 625 SCNALTSVTFPKSITTIGENAFDG 648  
 Identities = 43/134 (32%), Positives = 66/134 (49%), Gaps = 12/134 (8%)

Query: 511 MNRIGTLDLKDCLKIKIGDAAFHINHIYAIVLPESVQEIGRSAFRQNGALHLMFIGNKVK 570  
 + ++I T+ + D + +G+ AF + ++ LP S+ IG AF+ L + I N V  
 Sbjct: 66 LQSKIQTVTIGDGVTSVGNNAFSDCALTSVTLPSNLTAIGDHAFKGCGLTSITIPNSVT 125

Query: 571 TIGEMAFLS-NKLESVNLSEQQLKTIEVQAFSD-NALSEVVLPPNLQTIREEAFKRNHL 628  
 TIGE AF + L+S+ L L I A S L+ + +P ++ TI E AF  
 Sbjct: 126 TIGEWAFKGCGLKSITL--PNSLTAGQSALSGCTGLTSITIPNSVTTIGEWAF----- 178

Query: 629 KEVKGSSSTLSQITF 642  
 G S L+ ITF  
 Sbjct: 179 ---FGCSGLTSITF 189

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 801> which encodes the amino acid sequence <SEQ ID 802>. Analysis of this protein sequence reveals the following:

Possible site: 21  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -2.44 Transmembrane 984 -1000 ( 984 -1001)

----- Final Results -----  
 bacterial membrane --- Certainty=0.1977(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 975-979

An alignment of the GAS and GBS proteins is shown below:

Identities = 751/1050 (71%), Positives = 861/1050 (81%), Gaps = 45/1050 (4%)

Query: 3 KKHLKTLALALTTVSVVTYSQEVYGLEREESVKQEQTQSA-SEDDWFEEDNERKTNVSKE 61  
 KKHLKT+AL LTTVSVVT++QEV+ L +E +KQ Q S+ S D+ E + K +++  
 Sbjct: 2 KKHLKTVALTLTTVSVVTHNQEVFSLVKEPILKQTQASSISGADYAESSGKSKLKINET 61

Query: 62 NSTVDETVDLFDSDGNSNNSSSKTESVVSDDPKQVPAKPEVTQEASNSNDASKVEVPKQ 121  
 + VD+TV+DLFSD + K +Q KA E T E+ S++E K+  
 Sbjct: 62 SGFVDDTVTDLFDSDKRTTPEKIKDNLAKGPREQELKAVTENT-ESEKQITSGSQLEQSKE 120

-317-

5 Query: 122 DTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVLP SHAADGTQLTQVAS 181  
+ K TS WE DF+T+G+TLVG SKSG+ KLSQT HLVLP S AADGTQL QVAS  
Sbjct: 121 SLSLNKTVPSTSNWEICDFITKGNLTVLGSKSGVEKLSQTDHLVLP SQAADGTQLIQVAS 180

10 Query: 182 FAFTPDKKTAAIEYTSRLGENGKPSRLDIDQKEIIDEGEIFNAYQLTKLTIPNGYKSIGQ 241  
FAFTPDKKTAAIEYTSR GENG+ S+LD+D KEII+EGE+FN+Y L K+TIP GYK IGQ  
Sbjct: 181 FAFTPDKKTAAIEYTSRAGENGESQLDQDVGKEIINEGEVFN SYLLKKVTTIPTGYKHIGQ 240

15 Query: 242 DAFVDNKNIAEVLNLPESLETISDYAFAHMSLKQVKLPDNLKVIGELAFFDNQIGGKLYLP 301  
DAFVDNKNIAEVLNLPESLETISDYAFAH++LKQ+ LPDNLK IGELAFFDNQI GKL LP  
Sbjct: 241 DAFVDNKNIAEVLNLPESLETISDYAF AHLALQIDLPDNLK AIGELAFFDNQITGKLSLP 300

20 Query: 302 RHLIKLAERAFKSNRIQTVEFLGSKLVIGEASFQDNRLRNVM L PDGLEKIESEAF TGNP 361  
R L++LAERAFKSN I+T+EF G+ LKVIGEASFQDN+L +MLPDGLEKIESEAF TGNP  
Sbjct: 301 RQLMRLAERAFKSNHIKTIEFRGNSLVIGEASFQDN DLSQLMLPDGLEKIESEAF TGNP 360

25 Query: 362 GDEHYNNQVVLRTTRTQGNPHQLATENTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNSVT 421  
GD+HYNN+VVL T++G+NP LATENTYVNPDKSLW+ +P++DYTKWLEEDFTYQKNSVT  
Sbjct: 361 GDDHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420

30 Query: 422 GFSNKGKLVKRRNKNLEIPKQHNGITITEIGDNAFRNVDFQSKTLRKYDLEEIKLPSTIR 481  
GFSNKGKLVK+RNKNLEIPKQHNG+TITEIGDNAFRNVDFQ+KTLRKYDLEE+KLPSTIR  
Sbjct: 421 GFSNKGKLVKRRNKNLEIPKQHNGVTTITEIGDNAFRNVDFQNKTLRKYDLEEVKLPSTIR 480

35 Query: 482 KIGAFAFQSNNLKSFEASEDLEEIKEGAFMNNRIGTDLKDKLIKIGDAAFHINHIAIV 541  
KIGAFAFQSNNLKSFEAS+DLEEIKEGAFMNNRI TL+LKDKL+ IGDAAFHINHIAIV  
Sbjct: 481 KIGAFAFQSNNLKSFEASDDLEEIKEGAFMNNRIETLELKD KLVTIGDAAFHINHIAIV 540

40 Query: 542 LPESVQEIGRSAFRQNGALHLMFIGNKVKTIGEMAFSLNKLESVNLSEQKQLKTIEVQAF 601  
LPESVQEIGRSAFRQNGA +L+F+G+KVKT+GEMAFSLN+LE ++LSEQKQL I VQAF  
Sbjct: 541 LPESVQEIGRSAFRQNGANNLIFMGSKVKT LGEMAFSLNRLEHLDLSEQKQLTEIPVQAF 600

45 Query: 602 SDNALSEVVLPPNLQTIREEAFKRNHLKEVKGSSSTLSQITFNAFDQNDGDKRFGKKVVR 661  
SDNAL EV+LP +L+TIREEAFK+NLK+++ +S LS I FNA D NDGD++F KVVV+  
Sbjct: 601 SDNALKEVLLPASLKTIREEAFKRNHLKQLEVASALSHIAFNALDDNDGDEQFDNKVVVK 660

50 Query: 662 THNNSHMLADGERFIIDPKLSSTMVDLEKVLKIEGLDYSTLRQTTQTQFRMTTAGKA 721  
TH+NS+ LADGE FI+DPDKLSST+VDLEK+LK+IEGLDYSTLRQTTQTQFR+MTTAGKA  
Sbjct: 661 THNSYALADGEHFIVDPDKLSSTIVDLEKILKIEGLDYSTLRQTTQTQFRDMTTAGKA 720

55 Query: 722 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKA EKALVT KATKNGHLLERSINKAVL 781  
LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKA EKALVT KATKNG LLERSINKAVL  
Sbjct: 721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKA EKALVT KATKNGQLLERSINKAVL 780

60 Query: 782 AYNNSAIKKANVKRLEKELDLLTDLVEGKGPLAQATMVQGVYLLKTPPLPEYYIGLNVY 841  
AYNNSAIKKANVKRLEKELDLLT LVEGKGPLAQATMVQGVYLLKTPPLPEYYIGLNVY  
Sbjct: 781 AYNNSAIKKANVKRLEKELDLLTGLVEGKGPLAQATMVQGVYLLKTPPLPEYYIGLNVY 840

65 Query: 842 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVEDNEGYHTLAVATLADYEGLYIKDIILN 901  
FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVEDNEGYH LAVATLADYEG L IK ILN  
Sbjct: 841 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVEDNEGYHALAVATLADYEGLDIKTIILN 900

Query: 902 SSLDKIKAIRQIPLAKYHRLGIFQAIRNAAAEADRLLPKTPKGYLNEVPNYRKKQVEKNL 961  
S L ++ +IRQ+P A YHR GIFQAI+NAAAEA++LLPK  
Sbjct: 901 SKLSQLTSIRQVPTAAYHRAGIFQAIQNAAAEAEQLLPK----- 939

Query: 962 KPVDYKTPIFNKALPNEKVDGDRAAKGHNINAETNNSVAVTPIRSEQQLHKSQSDVNLPO 1021  
++++ + N++ ++S + ++ + LP+  
Sbjct: 940 -----PGTHSEKSSSESANSKDRG-----LQSNPKTNRGRHSAILPR 977

Query: 1022 TSSKNFIYEILGYVSLCLLFLVTAGKKGK 1051  
T SK +F+Y ILGY S+ LL L+TA KK K  
Sbjct: 978 TSGKSGFVYGILGYTSVALLSLITAIIKKK 1007

SEQ ID 800 (GBS97) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 12; MW 113.4kDa).

GBS97-His was purified as shown in Figure 193, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 254

- 5 A DNA sequence (GBSx0269) was identified in *S.agalactiae* <SEQ ID 803> which encodes the amino acid sequence <SEQ ID 804>. This protein is predicted to be ribonucleoside-diphosphate reductase alpha chain (nrdE). Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4274 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB96160 GB:AE000050 ribonucleoside-diphosphate reductase alpha  
chain-MPN324 (new), 513 (Himmelreich et al., 1996)  
[Mycoplasma pneumoniae]

Identities = 476/725 (65%), Positives = 586/725 (80%), Gaps = 20/725 (2%)

Query: 2 TQSD--AYLSINAKTRFRDRGTGNYHFTSDKEAVEQYMIEHVEPNTMVFTSLIEKLDYLVS 59  
TQ D +Y+SLNA T+ F D AVE Y+ EHV+P T VF S E+LD+LV

Sbjct: 12 TQEDLESYISLNAYTKVYG-----DFKMDLHAVEAYIQEHVKPKTKVVFHSTKERLDPLVK 66

Query: 60 NNYYESDLLKQYNLEFICQIFEHAYAKKFAFLNFMGALKFYNAVALKTEDNRYYLEHYED 119  
N+YY+ +++ Y+ E +I AYA +F + NFMGA KFYNAVALKT D ++YLE+YED

Sbjct: 67 NDYYDENIINMYSFEQFEEITRKAYAYRFRYANFMGAFKFYNAVALKTFDGKWYLENYED 126

Query: 120 RVVMNALFLAAGDEKAAYDLVDDMLANRFQPATPTFLNAGKKRGEYISCYLLRIEDNME 179  
RVVM LFLA G+ A L+ ++ NRFQPATPTFLNAG+K+RGE++SCYLLRIEDNME

Sbjct: 127 RVVMNVFLANGNYNKALKLLKQIITNRFQPATPTFLNAGKKRGEFVSCYLLRIEDNME 186

Query: 180 SISRAISTSLQLSKRGGGVALCLTNLREFGAPIKGIKNQATGIVPVMKLLDSFSYANQL 239  
SI RAI+T+LQLSKR GGVAL LTN+RE GAPIK I+NQ++GI+P+MKLLDSFSYANQL

Sbjct: 187 SIGRAITTTLQLSKRDGGVALLLTNIRESGAPIKKIENQSSGIIPIMKLLDSFSYANQL 246

Query: 240 GQRQGAGAVYLHAHHPEVLTFDLTKRENADEKIRIKSLSLGLVIPDITTFELAKANKDMAL 299  
GQRQGAGAVYLHAHP+V+ FLDTKRENADEKIRIKSLSLGLVIPDITF LAK N++MAL

Sbjct: 247 GQRQGAGAVYLHAHPDVMQFLDTKRENADEKIRIKSLSLGLVIPDITFTLAKNNEEMAL 306

Query: 300 FSPYDIERVYGKPMDSISITEEYETLLANADIRKTFISARKLFQTIAELHFESGYPIILF 359  
FSPYD+ YGKP+SDIS+TE Y LLAN I+KTFI+ARK FQT+AELHFESGYPIILF

Sbjct: 307 FSPYDVYEEYGKPLSDISVTEMYELLANQRIKKTFINARKFFQTV AELHFESGYPIILF 366

Query: 360 EDTVNAKNPHKKEGRIVMSNLCSEIAQVNTASQFSEDLTFTKVGHVCCNLGSINIARAM 419  
+DTVN +N H RIVMSNLCSEI Q +T S+F DL F KVG+D+ CNLGS+NIA+AM

Sbjct: 367 DDTVNRRNAH--PNRIVMSNLCSEIVQPSTPSEFHHDLAFKKVGNDISCNLGSINIARAM 424

Query: 420 DQAADEFKLIANSIRALDRVSRTSDLDAPSIIKKGNAANHAVGLGAMNLHGFLATNHIYY 479  
+ +F +L+ +I +LD VSR S+L++APSI+KGN+ NHA+GLGAMNLHGFLATN IYY

Sbjct: 425 ESGPEFSELVKLAIESLDLVSRSNLETAPSIQKGNSENHALGLGAMNLHGFLATNQIYY 484

Query: 480 DSQEADFTDCFFYAMAYYAFKASNLHAKKEGTFEGFSESSYADGSYFYQY--TEQNF-E 536  
+S EAIDFT+ FFY +AY+AFKAS+ LA EKG F+ F + +ADGSYF +Y E +F

Sbjct: 485 NSPEADFTNIFFTYVAYHAFKASSELAEKKGKFNFNENTKFADGSYFDKYIKVEPDFWT 544

Query: 537 PKTQRVKNLLAEYGLTLPSQEDWRKLVQSIKEIGLANAHLLAVAPTGSISYLSSTPSLQ 596  
PKT+RVK L +Y + +P++E+W++L +I++ GLAN+HLLA+APTGSISYLSSTPSLQ

Sbjct: 545 PKTERVKALFQKYQVEIPTRENWKEALNIQKNGLANSHLLAIAPTGSISYLSSTPSLQ 604



-319-

Query: 597 PVVSPVEVRKEGALGRVYPAYKIDADNYVYKKGAYEVGSEAIINIAAAQKHIDQAIS 656  
 PVVSPVEVRKEG LGR+YVPAY+++ D+Y +YK GAYE+G E IINIAAAQ+H+DQAIS  
 Sbjct: 605 PVVSPVEVRKEGRLGRIYVPAYQLNKDSYPFYKDGAYELGPEPIINIAAAQQHVDQAIS 664

Query: 657 LTLFMTDQATTRDLNKAYIQAFKQKCSIYYVRVRQDILEGSESYDDMLDDFTSSDLEDC 716  
 LTLFMTD+ATTRDLNKAYI AFK+ C+SIYYVRVRQ++LE SE + + ++ C  
 Sbjct: 665 LTLFMTDKATTRDLNKAYIYAFKKGCSIIYYVRVRQEVLEDSSEDH-----TIQMQQC 716

Query: 717 QSCMI 721  
 ++C+I  
 Sbjct: 717 EACVI 721

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 805> which encodes the amino acid  
 sequence <SEQ ID 806>. Analysis of this protein sequence reveals the following:

Possible site: 47  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1843(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC82625 GB:AF054892 surface antigen BspA [Bacteroides forsythus]  
 Identities = 124/451 (27%), Positives = 202/451 (44%), Gaps = 65/451 (14%)

Query: 221 FNSYLLKKVTIPTGYKHIGQDAFVDNKNAIEVNLPESETISDYAFAHLA-LKQIDLDPN 279

F+ L VT+P IG AF + + +P S+ TI ++AF + LK I LP++  
 Sbjct: 87 FSDCALTSVTLPNLSLTAIGDHAFKGCGLTSITIPNSVTTIGEWAFKGCGLKSITLPS 146

Query: 280 LKAIGELAFFDNQITGKLSLPRQLMRLAERA-FKSNHIKTIEFRGNSLKVIGEASFQD-N 337  
 L AIG+ A +++P + + E A F + + +I F NSL IGE++F  
 Sbjct: 147 LTAIGQSALSGCTGLTSITIPNSVTTIGEWAFGCSGLTSITF-PNSLTAIGESAFYGC 205

Query: 338 DLSQLMLPDGLEKIESEFTGNPGDDHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQ 397  
 L+ + LPD L I AF G G KS P+ L T +S +  
 Sbjct: 206 ALTSITLPDALTTIGESAFKGCSCG-----LKSITFPNSLTITIG-----ESAFY 248

Query: 398 ESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIPKQHNGVTITEIGDNAFRN 457  
 + + + T +++ G S GL K++ P ++T IG++AF N  
 Sbjct: 249 DCGALTSITLPDALTTIGRSFYGCS--GL-----KSITFPN-----SLTTIGESAFYN 295

Query: 458 VDFQNKTLRKYDLEEVKLPSTIRKIGAFAFQS--NNLKSFEASDDLEEIKEGAFMNNRIET 516  
 L + +P+++ IG AF + LKS D L I+E AF N + T  
 Sbjct: 296 CG-----SLTSITIPNSVTTIGRSFYGCSGLKSITLPDGLTTIEERAFYNCGLVT 346

Query: 517 -LELKDKLVTIGDAAFH-INHIYAIVLPESVQEIGRSAPFRQNGANNLIFMGSKVKTIGEM 574  
 + + + + TIG++AF+ + + +I LP+ + I AF GA I + + V T+GE  
 Sbjct: 347 SITIPNSVATIGESAFYGCGLKSITLPDGLTTIEWGAFYNCGALTSITIPNSVSTIGES 406

Query: 575 AFLS-NRLEHLDLSEQQLTEIPVQAFSDNALKEVLL--PASLKTIREEAFKKNHLKQLE 631  
 AF L+ + ++ + +I F + L + L PA KT+ E K+ K+  
 Sbjct: 407 AFYGCGLKDVTVAWDTPI-DIQRDVFRELTLSGIRLHVPAGKKTVE--AKDVWKE-- 460

Query: 632 VASALSHIAFNALDDND-GDEQFDNKVVVKT 661  
 FN ++D+D G Q++ KT  
 Sbjct: 461 -----FNIVEDDDFGGLQWNYDAATKT 482

An alignment of the GAS and GBS proteins is shown below:

Identities = 534/726 (73%), Positives = 614/726 (84%), Gaps = 5/726 (0%)

Query: 1 MTQSDA-YLSLNAKTRFRDRTGNYHFTSDKEAVEQYMIEHVEPNTMVFTSLIEKLDYLV 59

-320-

M+Q++A YLSLNA TRF+ G+YHF SDKEAV +Y+ EHV PN M F SL +KL YL++  
 Sbjct: 1 MSQTNASYSLSLNALTRFKKPDGSHFSDKEAVRRYLEEHVSPNQMAFNSLEDKLAYLIN 60

5 Query: 60 NNYYESDLLKQYNLEFICQIFEHAYAKKFAFLNFMGALKFYNAVALKTEDNRYYLEHYED 119  
 YYE + Y + I + F +AY + + FLN MGA+Kfy +YALKT D + YLE +ED  
 Sbjct: 61 EGYEQAIFDAYPNDLIKEAFHYAQGYRFLNLMGAMKfyQSYALKTLTGKQYLETfED 120

10 Query: 120 RVMNALFLAAGDEKAAYDLVDDMLANRFQPATPTFLNAGKKRRGEYISCYLLRIEDNME 179  
 R VMNALFLA GD+ +D++D +L RFQPATPTFLNAGKKRRGEYISCYLLR+EDNME  
 Sbjct: 121 RAVMNALFLADGDQTFVFDVIDAILHRRFQPATPTFLNAGKKRRGEYISCYLLRVEDNME 180

15 Query: 180 SISRAISTSLQLSKRGGVALCLTNLREFGAPIKGIKNQATGIVPVMKLLSDFSANQL 239  
 SISRAISTSLQLSKRGGVALCLTNLRE GAPIKGI+NQATGIVPVMKLLSDFSANQL  
 Sbjct: 181 SISRAISTSLQLSKRGGVALCLTNLREIGAPIKGIENQATGIVPVMKLLSDFSANQL 240

20 Query: 240 GQRQAGAGAVYLHAHHPEVLTFLDTKRENADEKIRIKSLGLVIPDITFELAKANKDMAL 299  
 GQRQAGAGAVYLHAHHPEVLTFLDTKRENADEKIRIKSL+LGLVIPDITF+LAK NKDMAL  
 Sbjct: 241 GQRQAGAGAVYLHAHHPEVLTFLDTKRENADEKIRIKSLALGLVIPDITFQLAKENKDMAL 300

25 Query: 300 FSPYDIERVYKPMDSISITEEYETLLANADIRKTFISARKLFQTIAELHFESGYPYILF 359  
 FSPYDI+R YGK MSDISITEEY+ LLAN I+KT+ISARK FQ IAEHFESGYPY+LF  
 Sbjct: 301 FSPYDIKRAYGKMDSDISITEEYDKLLANPAIKKTYISARKFFQLIAELHFESGYPYILF 360

30 Query: 360 EDTVNAKNPHKKEGRIVMSNLCSEIAQVNTASQFSEDLTFTKVGHVCCNLGSINIARAM 419  
 +DTVN +NPH K+GRIVMSNLCSEIAQV+T S F EDL+F +G D+CCNLGSINIA+AM  
 Sbjct: 361 DDTVNKNRNPAAKGRIVMSNLCSEIAQVSTPSTFKEDLSFETIGEDICCNLGSINIAQAM 420

35 Query: 420 DQAADFELIANIRALDRVSRSDLSAPSIIKGNANHAVGLGAMNLHGFLATNHIYY 479  
 A FE+LI SIRALDRVSR SDL+ APS++ GNAANHAVGLGAMNLHGFLATNHIYY  
 Sbjct: 421 ADAPHFEQLITTSIRALDRVSRVSDLCAPSIVETGNAANHAVGLGAMNLHGFLATNHIYY 480

40 Query: 480 DSQEAIDFTDCFFYAMAYYAFKASNHLAKEKGTfEGFSESSYADGSYFYQYTEQNFEPKT 539  
 D++EA+DFTD FF+AMAYYAFKAS LAKEKG F GFS S+Y+DG+YF +Y +++ +P+T  
 Sbjct: 481 DTKEAVDFTDLFFHAMAYYAFKASCQLAKEKGAFAGFSLSTYSdGTYFAKYLQEDAKPQT 540

45 Query: 540 QRVKNLLAEYGLTLPSQEDWRKLVQSIKEIGLANAHLAVAPTGSISYLSSTPQLPQV 599  
 +V LL +YG TLP+ DW+ LV IK+ GLANAHLAVAPTGSISYLSSTPQLPQV  
 Sbjct: 541 AKVATLLQDYGFTLPTVADWQALVADIKQFGLANAHLAVAPTGSISYLSSTPQLPQV 600

50 Query: 600 SPVEVRKEGALGRVYPAYKIDADNYVYKKGAYEVGSEAIINIAAAQKHIDQAISLTL 659  
 +PVEVRKEG+LGR+YVPAY+ID NY YY++GAYEVG +AII++ AAAQKH+DQAISLTL  
 Sbjct: 601 APVEVRKEGSLGRIYVPAYQIDQANYAYYERGAYEVGPKAIIIDVVAQAQKHVDQAISLTL 660

Query: 660 FMTDQATTRDLNKAYIQAFKQKASIYVVRVQDILEGSESYDD----MLDDFTSSDLED 715  
 FMTDQATTRDLN++YIQAFKQ CASIYVVRVQD+L GSE YD+ + +  
 Sbjct: 661 FMTDQATTRDLNRSYIQAFKQNCASIYVVRVQDVLGSEQYDEDSLVTAPGASDETTTE 720

Query: 716 CQSCMI 721  
 CQSCMI  
 Sbjct: 721 CQSCMI 726

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 255

55 A DNA sequence (GBSx0270) was identified in *S.agalactiae* <SEQ ID 807> which encodes the amino acid sequence <SEQ ID 808>. This protein is predicted to be nrdI protein (nrdI). Analysis of this protein sequence reveals the following:

Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2952(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-321-

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database:

5 >GP:AAC71451 GB:U39702 nrdI protein (nrdI) [Mycoplasma genitalium]  
 Identities = 77/127 (60%), Positives = 104/127 (81%), Gaps = 1/127 (0%)

Query: 7 VVYFSSKSNNTHRFVQKLACSNQRIPSD-GSSILVTEDYILIVPTYAGGGDDTKGAVPKQ 65  
 +VYFSS SNNTHRF++KL ++RIP D SI V+ +Y+LI PTY+GGG+ +GAVPKQ  
 10 Sbjct: 22 IVYFSSISNNTHRFIEKLGFGQHKRIPVDITQSI TVSNEYVLICPTYSGGQNQVEGAVPKQ 81

Query: 66 VVQFLNVRQNREHCQGVISSGNTNFGDTYAIAGPIIARKLNVP LLHQFELLGTQEDVTRV 125  
 V+QFLN + NRE C+GVI+SGNTNFGDT+ +AG +I++KLNVP LL+QFELLGT+ DV +  
 Sbjct: 82 VIQFLNNKHNRCLCRGVIASGNTNFGDTFCIAGTVISKLNVP LLYQFELLGT KNDVEQT 141

15 Query: 126 KELLCOF 132  
 +++++ F  
 Sbjct: 142 QKIIANF 148

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 809> which encodes the amino acid  
 sequence <SEQ ID 810>. Analysis of this protein sequence reveals the following:

Possible site: 54  
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.0089(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

30 Identities = 84/125 (67%), Positives = 100/125 (79%)

Query: 7 VVYFSSKSNNTHRFVQKLACSNQRIPSDGSSILVTEDYILIVPTYAGGGDDTKGAVPKQV 66  
 +VYFSSKSNNTHRFVQKL QRIP D + V+ Y+LIVPTYA GG D KGAV KQV  
 35 Sbjct: 6 IVYFSSKSNNTHRFVQKLGLPAQRIPVDNRPLEVSTHYLLIVPTYAAGGSDAKGAVSKQV 65

Query: 67 VQFLNVRQNREHCQGVISSGNTNFGDTYAIAGPIIARKLNVP LLHQFELLGTQEDVTRVK 126  
 ++FLN NR+HC+GVISSGNTNFGDT+A+AGPII++KL VPLLHQFELLGT DV +V+  
 Sbjct: 66 IRFLNPNNRKHCKGVISSGNTNFGDTFALAGPIISQKLQVPLLHQFELLGTATDVKKVQ 125

40 Query: 127 ELLCQ 131  
 + +  
 Sbjct: 126 AIFAR 130

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 vaccines or diagnostics.

**Example 256**

A DNA sequence (GBSx0271) was identified in *S.agalactiae* <SEQ ID 811> which encodes the amino acid  
 sequence <SEQ ID 812>. This protein is predicted to be ribonucleoside-diphosphate reductase beta chain  
 (nrdF). Analysis of this protein sequence reveals the following:

50 Possible site: 27  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 55 bacterial cytoplasm --- Certainty=0.3889(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

-322-

>GP:AAB96162 GB:AE000050 ribonucleoside-diphosphate reductase beta  
chain [Mycoplasma pneumoniae]  
Identities = 261/335 (77%), Positives = 301/335 (88%)

5 Query: 2 QSY YDRSQSP LDYALSEKAFPMRSVNWKNLNDKDL E VWN RV TQNF WLPEKIPVSNDLNS 61  
+ Y+ S SPL+YA + +RSVNW +D+KDLEVWNR+TQNF WLPEKIPVSND+ S  
Sbjct: 5 KKYFLESVSPLE YAQKKFQGNLRSVNWNLVDDEKDL E VWN RITQNF WLPEKIPVSNDIPS 64

10 Query: 62 WRTLDADWQQLITRTFTGLTLLDSVQATVGDIAQIKHSQTDHEQVIYANFAFMVAIHARS 121  
W+ L +WQ LIT+TFTGLTLLD++QAT+GDI QI ++ TDHEQVIYANFAFMV +HARS  
Sbjct: 65 WKQLSKEWQDLITKTFTGLTLLDTIQATIGDIKQIDYALTDHEQVIYANFAFMVGVHARS 124

15 Query: 122 YGTIFSTLCTSQIEEAHEWVVDTESLQARSRLIPFYTGDDPLKSKVAAAMMPGFLLYG 181  
YGTIFSTLCT S+QI EAHEWVW TESLQ R++ LIP+YTG DPLKSKVAAA+MPGFLLYG  
Sbjct: 125 YGTIFSTLCTSEQITEAHEWVVKTESLQKRAKALIPYYTGKDPLKSKVAAALMPGFLLYG 184

20 Query: 182 GFYLPFYLSARGKLPNTSDIIRLILRDKVIHNYYSYGKYQQKVAKLSVEKQAEMKTFVFD 241  
GFYLPFYLS+R +LPNTSDIIRLILRDKVIHNYYSYGKY+Q+KV K+S EKQAEMK FVFD  
Sbjct: 185 GFYLPFYLSRKQLPNTSDIIRLILRDKVIHNYYSYGKYFQRKVEKMSKEKQAEMKRFVFD 244

25 Query: 242 LLYQLIDLEKAYLYELYDGFDAEDAIRFSIYNAGKFLQNLGYDSPFTEETRISPEVFA 301  
L+Y+LI+LEKAYL ELY+GF + EDAI+FSIYNAGKFLQNLGYDSPFTEETRI PE+FA  
Sbjct: 245 LMYELIELEKAYLKELYEGFGIVEDAIKFSIYNAGKFLQNLGYDSPFTEETRIKPEIFA 304

30 Query: 302 QLSARADENHDFSGNGSSYIMGITEETLDEDWEF 336  
QLSARADENHDFSGNGSSY+MGI+EET D+DW+F  
Sbjct: 305 QLSARADENHDFSGNGSSYVMGISEETEDKDWDF 339

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 813> which encodes the amino acid  
30 sequence <SEQ ID 814>. Analysis of this protein sequence reveals the following:

Possible site: 18  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.3779(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

40 Identities = 292/335 (87%), Positives = 318/335 (94%)

Query: 2 QSY YDRSQSP LDYALSEKAFPMRSVNWKNLNDKDL E VWN RV TQNF WLPEKIPVSNDLNS 61  
Q YY+RSQSP++YALSE +RS+WNW LNDDKDL E VWN RV TQNF WLPEK+PVSNDLNS  
45 Sbjct: 3 QHY YERSQSP IEYALSETQKQLRSINWNYLNDDKDL E VWN RV TQNF WLPEKVPVSNDLNS 62

Query: 62 WRTLDADWQQLITRTFTGLTLLDSVQATVGDIAQIKHSQTDHEQVIYANFAFMVAIHARS 121  
WR+L DWQQLITRT+TGLTLLD+VQATVGD+AQI+HSQTDHEQVIY NFAFMV IHARS  
Sbjct: 63 WRS LGEDWQQLITRTYTGLTLLDTVQATVGDVAQIQHSQTDHEQVIYTNFAFMVGIHARS 122

50 Query: 122 YGTIFSTLCTSQIEEAHEWVVDTESLQARSRLIPFYTGDDPLKSKVAAAMMPGFLLYG 181  
YGTIFSTLC+S+QIEEAHEWVW T+SLQ R+R+LIP+YTGDDPLKSKVAAAMMPGFLLYG  
Sbjct: 123 YGTIFSTLCSSEQIEEAHEWVSTQSLQDRARVLIPIYTGDDPLKSKVAAAMMPGFLLYG 182

55 Query: 182 GFYLPFYLSARGKLPNTSDIIRLILRDKVIHNYYSYGKYQQKVAKLSVEKQAEMKTFVFD 241  
GFYLPFYLSARGK+PNTSDIIRLILRDKVIHNYYSYGKYQQKVA+LS EKQAEMK FVFD  
Sbjct: 183 GFYLPFYLSARGKMPNTSDIIRLILRDKVIHNYYSYGKYQQKVARLSPEKQAEMKAFVFD 242

60 Query: 242 LLYQLIDLEKAYLYELYDGFDAEDAIRFSIYNAGKFLQNLGYDSPFTEETRISPEVFA 301  
LLY+LIDLEKAYL ELY GFDLAEDAIRFS+YNAGKFLQNLGY+SPFT+EETR+SPEVFA  
Sbjct: 243 LLYELIDLEKAYLRELYAGFDLAEDAIRFSLYNAGKFLQNLGYSPFTDEETRVSPVFA 302

Query: 302 QLSARADENHDFSGNGSSYIMGITEETLDEDWEF 336  
QLSARADENHDFSGNGSSY+MGITEET D+DWEF  
Sbjct: 303 QLSARADENHDFSGNGSSYVMGITEETDDDDWEF 337

65

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 257

A DNA sequence (GBSx0272) was identified in *S.agalactiae* <SEQ ID 815> which encodes the amino acid sequence <SEQ ID 816>. This protein is predicted to be rhamnosyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 55  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1741(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9583> which encodes amino acid sequence <SEQ ID 9584> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA32090 GB:AB010970 rhamnosyltransferase [Streptococcus mutans]  
Identities = 104/309 (33%), Positives = 173/309 (55%), Gaps = 21/309 (6%)

Query: 11 QINICLATYNGQKYLRQQLDSIIQQGYTDWICLIRDDGSTDDTVAIKEYVNRDSRFIFI 70  
++NI ++TYNGQ+++ QQ+ SI +Q + +W LIRDDGS+D T II ++ D+R FI  
Sbjct: 2 KVNILMSTYNGQEFIAQQIQSIQKQTFENWNLLIRDDGSSDGTPIADFAKSDARIRFI 61

Query: 71 NSNDDRKLGSRSFYELVNYKKADFYVFSDDQDDVWKENRLERYLEEAKEFNQELPLLVYS 130  
N++ G ++FY L+ Y+KAD+Y FSDQDDVW +LE L EK N ++PL+VY+  
Sbjct: 62 NADKRENFVGIKNFYTLKYEKADYFFSDDQDDVWLPQKLELTASVEKENNQIPLMVYT 121

Query: 131 NWTSDVEKLTVL-----KEHNPATVIEQIAFNQINGMVIMMNHAKLWE--YRQIG 181  
+ T VD L VL + H+ T + E++ N + G +M+NH LAK W+ Y +  
Sbjct: 122 DLTVDVDRDLQVLHDSMIKTQSHHANTSLLLELTENTVTGGTMMVNHCLAKQWKQCYDDLI 181

Query: 182 AHD SYVGTLAYAVGNVAYISDSTVLWRRQ---VGAES---LNNYGRQYG-VATFWQMI 232  
HD Y+ LA ++G + Y+ ++T L+R+ +GA + L N+ R + V +W ++  
Sbjct: 182 MHDWYLALLAASLGKLIYLDETTELYRQHESNVLGARTWSKRLKNWLRPHRLVKKYWWLV 241

Query: 233 NTSFDRASLIFAQVSDKMSLERKLFSSRFIELKNANLMRRIYLLSKLKLRRKSLKETVAM 292  
+S +AS + + + K ++ L + + + RI L + + T  
Sbjct: 242 TSSQQQASHL---LELDLPAANKAIIRAYVTLLDQSFLNRIKWLKQYGFKNRAFHTFVF 298

Query: 293 TILLLTGYG 301  
L++T +G  
Sbjct: 299 KTLIITKFG 307

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 258

A DNA sequence (GBSx0273) was identified in *S.agalactiae* <SEQ ID 819> which encodes the amino acid sequence <SEQ ID 820>. Analysis of this protein sequence reveals the following:

Possible site: 36  
>>> Seems to have no N-terminal signal sequence  
INTEGRAL Likelihood = -4.19 Transmembrane 1213 -1229 (1211 -1230)

----- Final Results -----

bacterial membrane --- Certainty=0.2678(Affirmative) < succ>

-324-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9581> which encodes amino acid sequence <SEQ ID 9582> was also identified.

There is also homology to SEQ ID 822.

A related GBS gene <SEQ ID 8525> and protein <SEQ ID 8526> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1    Crend: 7  
 SRCFLG: 0  
 McG: Length of UR:    3  
       Peak Value of UR:    2.28  
       Net Charge of CR:    4  
 McG: Discrim Score:    1.29  
 GvH: Signal Score (-7.5): 2.84  
       Possible site: 30  
 >>> Seems to have a cleavable N-term signal seq.  
 Amino Acid Composition: calculated from 31  
 ALOM program    count: 0 value:    1.16 threshold:    0.0  
       PERIPHERAL    Likelihood =    1.16        344  
       modified ALOM score:    -0.73  
 \*\*\* Reasoning Step: 3  
 ----- Final Results -----  
       bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
       bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
       bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>  
 LPXTG motif: 1197-1201

SEQ ID 8526 (GBS147) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 4; MW 132kDa).

The GBS147-His fusion product was purified (Figure 200, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 286), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 259

A DNA sequence (GBSx0274) was identified in *S.agalactiae* <SEQ ID 823> which encodes the amino acid sequence <SEQ ID 824>. This protein is predicted to be Acetyltransferase (GNAT) family. Analysis of this protein sequence reveals the following:

Possible site: 57  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
       bacterial cytoplasm --- Certainty=0.2781(Affirmative) < succ>  
       bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
       bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG03505 GB:AE004449 conserved hypothetical protein [Pseudomonas aeruginosa]

-325-

Identities = 66/143 (46%), Positives = 94/143 (65%), Gaps = 5/143 (3%)

Query: 2 WNVKTFDNLTTHLFIQYKLRVSFVVEQDCPYQEVDDEDLI--CLHGMNWVDGQLAAYY 59  
 W K +LT EL+ + +LR VFVVEQ CPYQEV DL+ H M W DGQL AY  
 Sbjct: 5 WTCKHHADLTCLKELYALLQLRTEVFVVEQKCPYQEV DGLDLVGDTHHLMAWRDGQLLAYL 64

Query: 60 RLIP---EDDKVHLGRVIVNPDFRKKGLGNQLVEYAIFSEANYPNKPIYAQAQAYLQDF 116  
 RL+ + +V +GRV+ + R +GLG+QL+E A++ +E + + P+Y AQA+LQ +  
 Sbjct: 65 RLLEDVVRHEGQVVIGRVVSSSAARGQGLGHQLMERALQAAERLWLDTPVYLSAQAHLQAY 124

Query: 117 YQSFQFQPVSDIYLEDNIPHLDM 139  
 Y +GF V+++YLED+IPH+ M  
 Sbjct: 125 YGRYGFVAVTEVYLEDIPHIGM 147

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 260

20 A DNA sequence (GBSx0275) was identified in *S.agalactiae* <SEQ ID 825> which encodes the amino acid sequence <SEQ ID 826>. Analysis of this protein sequence reveals the following:

Possible site: 37  
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2010(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

30 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 261

35 A DNA sequence (GBSx0276) was identified in *S.agalactiae* <SEQ ID 827> which encodes the amino acid sequence <SEQ ID 828>. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2935(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

45 >GP:CAB12631 GB:Z99108 similar to RNA methyltransferase [Bacillus subtilis]  
 Identities = 217/448 (48%), Positives = 298/448 (66%), Gaps = 4/448 (0%)

Query: 7 QRIPLKIKRMGINGEGIGFYKKTILIFVPGALKGEEVFCQISSVRNFAEAKLLKINKKSK 66  
 Q PL IKR+GINGEG+G++KK ++FVPGAL GEEV Q + V+ F+E ++ KI K S+  
 50 Sbjct: 16 QTFLPTIKRLGINGEGVGYYKKKVVFPVPGALPGEEVVQATKVQPKFSEGRICKIRKASE 75

Query: 67 NRVEPPCSIYKECGGCQIMHLQYDKQLEFKTDVIRQALMKFKPEGYENYEIRKTIGMSEP 126  
 +RV PPC +Y++CGGCQ+ HL Y +QL K D++ Q+L + EN EI++TIGM P

-326-

Sbjct: 76 HRVAPPCPVYEQCGGCQLQHLAYSQQLEKRDIVIQSLERHTKFKVENMEIKETIGMDNP 135

Query: 127 EHYRAKLQFQV-RSFGGNVKAGLYAQGTHRLIDIKDCLVQDSLQEMINRVAELLGKYKL 185  
 +YR K QFQ+ RS G++ AGLY +H ++ IKDC+VQ T + V +L ++

5 Sbjct: 136 WNYRNSQFQIGRSQSGSIIAGLYGLSDHIVPIKDCIVQHPATNKTTGIVRRILEDNFV 195

Query: 186 PIYNERKIAG-VRTVMIRRAQASGEVQLIFITTSKRL--DFDDVVIELVREFPELKTIVAVN 242  
 +YNERK G VRT++ R +GEVQ++ +T+K +++V + + PE+K++ N

10 Sbjct: 196 SVYNERKRKGDVRTIVTRVGFETGEVQVVLVTAKETLPHKEEIVKAIQKRLPEVKSIIQN 255

Query: 243 INASKTSDIYGQITEVIWGQESINEEVLVDYGFSLSPRAFYLNPQTQIILYSEAVKALDV 302  
 +N +KTS I+G+ T+ + G+ I E + D F LS RAF+QLNP+QT LY E KA +

Sbjct: 256 VNGAKTSVIFGEKTKQLAGKTVIQEVLGDVSFELSARAFFQLNPEQTVKLYDEVKKAQQL 315

15 Query: 303 KEDDDDLIDAYCGVGTIGLAFAGKVKSVRGMDIPEAIQDAKENALYMGFTNTHYEAGKAE 362  
 + ++DAYCGVGTIG+ A K VRGMD+I E+I DAK+NA G N Y G AE

Sbjct: 316 TGKEKVVDDAYCGVGTIGMWVADGAKEVRGMDVIKESIDDAKNAKKGHMANATYVTGTAE 375

Query: 363 DIIPRWYSEGFRANALIVDPPTGLDDKLLNTILKMPPEKMVYVSCNTSTLARDLVTLTK 422  
 +P+W EGFR + +IVDPPRTG D L+TI K+ P++ VYVSCN STLA+DL TL+K

20 Sbjct: 376 HWLPKWTKEGFRPDVIVDPPTGCDSTFLDTIKKVKPKRFVYVSCNPSTLAKDLQTLTK 435

Query: 423 VYHVHYIQSVDMPHTARTEAVVKLQRK 450  
 Y V YIQ VDMFP TA EAV +L K

25 Sbjct: 436 DYRVDIQPVDMFPQTAEVAVARLVK 463

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 829> which encodes the amino acid sequence <SEQ ID 830>. Analysis of this protein sequence reveals the following:

Possible site: 56  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2980(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 327/450 (72%), Positives = 397/450 (87%)

40 Query: 1 MNVVLKQRIPLKIKRMGINGEGIGFYKKTILFVPGALKGEEVFCQISSVRRNF AEAKLLK 60  
 M V +KQ+IPLKIKRMGINGEGIGFY+KTL+FVPGALKGE++FCQI++V+RNFAEAKLL

Sbjct: 1 MVVKVKQKIPLKIKRMGINGEGIGFYQKTLVFPVPGALKGEDIFCQITAVKRNFAEAKLLT 60

45 Query: 61 INKSKNRVPEPPCSIYKECGGCQIMHLQYDKQLEFKTDVIRQALMKFKPEGYENYEIRKT 120  
 +NK SKNRV+P CS+Y+ CGGCQIMHL Y KQL+FK DVIRQAL KFKP GYE +EIR T

Sbjct: 61 VNKASKNRVKPACSVYETCGGCQIMHLAYPKQLDFKDDVIRQALKKFKPTGYEQFEIRPT 120

Query: 121 IGMSEPEHYRAKLQFQVRSFGGNVKAGLYAQGTHRLIDIKDCLVQDSLQEMINRVAELL 180  
 +GM +P+HYRAKLQFQ+RSFGG VKAGL++QG+HRL+ I +CLVQD LTQ++IN++ +L+

50 Sbjct: 121 LGMKKPDHYRAKLQFQLRSFGGTVKAGLFSQGSRLVPIDNCLVQDQLTQDIINKITQLV 180

Query: 181 GKYLPIYNERKIAGVRTVMIRRAQASGEVQLIFITTSKRLDFDDVVIELVREFPELKTIVA 240  
 KYKLPIYNERKIAG+RT+M+R+AQAS +VQ+I ++SK + + + EL + FP++KTVA

55 Sbjct: 181 DKYKLPIYNERKIAGIRTIMVRKAQASDQVQIIIVSSKEVRLANFIGELTKAFPQVKIVA 240

Query: 241 VNINASKTSDIYGQITEVIWGQESINEEVLVDYGFSLSPRAFYLNPQTQIILYSEAVKAL 300  
 +N N SK+S+IYG TE++WGQE+I+EEVLVDYGF+LSPRAFYLNP+QT++LY E VKAL

Sbjct: 241 LNSNRKSSEIYGDETEILWGQEAIEEVLVDYGFALS PRAFYLNPQQTTEVLYGEVVKAL 300

60 Query: 301 DVKEDDDLIDAYCGVGTIGLAFAGKVKSVRGMDIPEAIQDAKENALYMGFTNTHYEAGK 360  
 DV D +IDAYCGVG+IG AFAGKVKSVRGMDIPEAI+DA++NA MGF N +YEAGK

Sbjct: 301 DVGSKDHIIDAYCGVGSIGFAFAGKVKSVRGMDIPEAIEDAQKNAKAMGFDNAYYEAGK 360

65 Query: 361 AEDIIPRWYSEGFRANALIVDPPTGLDDKLLNTILKMPPEKMVYVSCNTSTLARDLVTL 420  
 AEDII +WY +G+RA+A+IVDPPRTGLDDKLL TIL P++MVYVSCNTSTLARDLV L



-327-

Sbjct: 361 AEDIISKWYKQGYRADAVIVDPRTGLDDKLLKTLILHYQPKQMVVVSNTSTLARDLVQL 420

Query: 421 TKVYHVHYIQSVDMFPHTARTEAVVKLQRK 450  
TKVY VHYIQSVDMFPHTARTEAVVKLQ++

5 Sbjct: 421 TKVYDVHYIQSVDMFPHTARTEAVVKLQKR 450

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 262

- 10 A DNA sequence (GBSx0277) was identified in *S.agalactiae* <SEQ ID 831> which encodes the amino acid sequence <SEQ ID 832>. Analysis of this protein sequence reveals the following:

Possible site: 45

>>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3505(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database:

>GP:BA04643 GB:AP001510 unknown conserved protein in B. subtilis  
[Bacillus halodurans]

Identities = 74/263 (28%), Positives = 141/263 (53%), Gaps = 9/263 (3%)

25 Query: 3 ITKIEKKKR---LYTLEL-DNTENLY---ITEDTIVHFMLSKGMIINAEKLENIKKFAQL 55  
IT+IE +KR Y + + N +++Y + E ++ L KG+ I+AE+++ I ++

Sbjct: 4 ITRIEVQKRNNERYNIFIHQNGQDVYAFSVDEQVLIQGLRKGLDIDAEQMKQILYEDEV 63

30 Query: 56 SYGKNLGLYYISFKQRTKEVEIKYLQQHDIDSKIIPQIIDNLKSENWINDKNYVQSFIQQ 115  
NL L+Y+S++ R+ EV YL++ D + II ++ L + ++D + ++FIQ

Sbjct: 64 QKTFNLALHYLSYRMRSVHEVRTYLKKKDRREEPIIEHVLHRLTEQRLDDHAFAEAFIQT 123

35 Query: 116 NLNTGDKGPYVIKQKLLQKGIKSKIIESELQAINFQDLASKISQKLYKKYQNKPLKAL- 174  
T KGP +KQ+L +KG+ K IE L ++++ ++ L K+ +L

Sbjct: 124 KRATTSGPLKQELAEKGVSEKTIIEGALTTFSEYEQVEQVKAWLEKQKGRFTFGSSLA 183

40 Query: 175 -KDKLMQSLTTKGFQIVHTVIONLEIEKDQLEEDLIYKELDKQYQKLSKKHDQYELK 233  
K KL + L KG+ ++ ++ I++++E E + + +K +K + K +EL+

Sbjct: 184 WKQKLSRQLLAKGYTSPVIEEAFADVPIKQEEEEWEALKAFGEKAMRKYAGKKTGWELQ 243

45 Query: 234 QRIINALMRKGYQYEDIKSALRE 256

Q++ AL RKG+ E I+ L +

Sbjct: 244 QKVKQALYRKGFSLEMIERYLND 266

- 45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 833> which encodes the amino acid sequence <SEQ ID 834>. Analysis of this protein sequence reveals the following:

Possible site: 58

>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2388(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 55 An alignment of the GAS and GBS proteins is shown below:

Identities = 146/258 (56%), Positives = 190/258 (73%)

Query: 1 MKITKIEKKRKYLYTLELDNTENLYITEDTIVHFMLSKGMIINAEKLENIKKFAQLSYGKN 60  
MKITKIEKKRKYLY +ELDN E+LY+TEDTIV FMLSK +++ ++LE++K FAQLSYGKN

-328-

Sbjct: 1 MKITKIEKKKRLYLIELDNDESLYVTEDTIVRFMLSKDKVLDNDQLEDMKHFAQLSYGKN 60

Query: 61 LGLYYISFKQRTKEVIKYLQQHDIDSKIIPQIIDNLKSENWINDKNYVQSFIQQNLNTG 120  
L LY++SF+QR+ K+V YL++H+I+ II II L+ E WI+D ++I+QN G

5 Sbjct: 61 LALYFLSFQQRSNKQVADYLRKHEIEEHIIADIITQLQEEQWIDDTKLADTYIRQNQLNG 120

Query: 121 DKGPIYIKQKLLQKGIKSKIIIESELQAINFQDLASKISQKLYKKYQNKLPKALKDKLMQ 180  
DKGP V+KQKLLQKGI S I+ L +F LA K+SQKL+ KYQ KLP KALKDK+ Q

10 Sbjct: 121 DKGPFVLKQKLLQKGIASHDIDPILSQTDFSQLAQKVSQKLFDKYQEKLPKALKDKKITQ 180

Query: 181 SLTTKGFYQIVHTVIQNLEIEKDQEELEDLIYKELDKQYQKLSKKHDQYELKQRIINAL 240  
+L TKGF Y + + +L ++D + EDL+ KELDKQY+KLS+K+D Y LKQ++ AL

15 Sbjct: 181 ALLTKGFSYDLAKHSLNHLNFDQDNQETEDLLDKELDKQYRKLRSKYDGYTLKQKLYQAL 240

Query: 241 MRKGQYEDIKSALREYL 258  
RKGY +DI LR YL

Sbjct: 241 YRKGYNSSDDINCKLRNYL 258

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 263

A DNA sequence (GBSx0278) was identified in *S.agalactiae* <SEQ ID 835> which encodes the amino acid sequence <SEQ ID 836>. Analysis of this protein sequence reveals the following:

Possible site: 33  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.3912(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04659 GB:AP001510 unknown conserved protein in B. subtilis  
[Bacillus halodurans]

35 Identities = 96/175 (54%), Positives = 122/175 (68%)

Query: 1 MRLPKEGDFITIQSYKHDGSLHRTWRDTMVLKTTENALIGVNDHTLVTENDGRRWVTREP 60  
M PK G I IQSYKH+GS+HR W +T+VLK T +IG ND LV E+DGR W TREP

40 Sbjct: 1 MNFPKVGSKIQIQSYKHNGSIHRIWEETIVLKGTSKVIGGNDRILVKESDGRHWRTREP 60

Query: 61 AIVYFHKKYWFNIIAMIRETGVSYYCNLASPYILDPEALKYIDYDLVDVKVFADGEKRLLD 120  
AI YF + WFN I MIR G+ +YCNL +P+ D EALKYIDYDL+KVF D +LLD

Sbjct: 61 AICYFDSEQWFNTIGMIRADGIYFYCNLGPFTWDEEALKYIDYDLDIKVFDPDMTFKLLD 120

45 Query: 121 VDEYEQHK AQMNYPTDIDYILKENVKILVEWINENKGPFSYINIWYKRYLELK 175  
DEY H+ M YP +ID IL+ +V LV WI++ KGPF+ ++ WY+R+L+ +

Sbjct: 121 EDEYAMHRKMMKYPPEIDRILQRSVDELVSWIHQKGFAPQFVSWYERFLQYR 175

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 837> which encodes the amino acid sequence <SEQ ID 838>. Analysis of this protein sequence reveals the following:

Possible site: 33  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.3912(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

-329-

Identities = 155/177 (87%), Positives = 165/177 (92%)

Query: 1 MRLPKEGDFITIQSYKHDGSLHRTWRDTMVLKTTENALIGVNDHTLVTE+DGRRWVTREP 60  
 M+LPKEGDFITIQSYKHDGSLHRTWRDTMVLKTTENALIGVNDHTLVTE+DGRRWVTREP  
 Sbjct: 1 MKLPKEGDFITIQSYKHDGSLHRTWRDTMVLKTTENALIGVNDHTLVTESDGRRWVTREP 60

Query: 61 AIVYFHKKYWFNIIAMIRETGVSYYCNLASPYILDPEALKYIDYDLVDVKVFADGEKRLLD 120  
 AIVYFHKKYWFNIIAMIR+ GVSYYCNLASPY++D EALKYIDYDLVDVKVFADGEKRLLD  
 Sbjct: 61 AIVYFHKKYWFNIIAMIRDNGVSYYCNLASPYMDTEALKYIDYDLVDVKVFADGEKRLLD 120

Query: 121 VDEYEQHKAQMNYPTDIDYILKENVKILVEWINENKGPFSYINIWKRYLELKKR 177  
 VDEYE HK +M Y D+D+ILKENVKILV+WIN KGPF +YI IWYKRYLELK R  
 Sbjct: 121 VDEYETHKKEMQYSADMDFILKENVKILVDWINHEKGPFSKAYITIWYKRYLELKNR 177

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 264

A DNA sequence (GBSx0288) was identified in *S.agalactiae* <SEQ ID 839> which encodes the amino acid sequence <SEQ ID 840>. This protein is predicted to be jag protein. Analysis of this protein sequence reveals the following:

Possible site: 49  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 25 bacterial cytoplasm --- Certainty=0.1666(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

- 30 >GP:BAB07782 GB:AP001520 spoIIJ-associated protein [Bacillus halodurans]  
 Identities = 54/198 (27%), Positives = 98/198 (49%), Gaps = 6/198 (3%)
- Query: 100 DVVEEYIEEVDETLEKEDVSQPELPKIDDKNVVTTSEAEKIDLLPNIEVAAAQVTKYVE 159  
 + VE+ I E+ T E+ + E PK ++ + A+ ++ + P+ + ++E  
 35 Sbjct: 13 EAVEQAIIELGTTREIRITYTVVEEPKSGLFILGSKPAVIEVVVKPD---PVDRAKAFLE 69
- Query: 160 NIIYEMDLDA--TIETTTSKRQINLQIETPEAGRIIGYHGKVLKSLQLLAQNYLHDFRSK 217  
 ++ EMD++ TIE + N+ E + G +IG G+ L SLQ L + +  
 40 Sbjct: 70 ELLQEMDMEVEVTIEKDPATVLFNISGEQ-DLGLTIGKRGQTLDLSQYLVNVLVANKEEGE 128
- Query: 218 SFSVSINVHDYVEHRTETLIDFSKKIARRVLETNEPYHMDPMSNSERKTVHKTITATIEGV 277  
 + ++ +Y R E L+ +++A + L T P ++PMS ERK +H + + V  
 45 Sbjct: 129 FIRIKLDAENYRARRKEALVQLAERLASKALRTKRPVSLEPMSAHERKIHTALQELGDV 188
- Query: 278 ESYSEGNDPNRFVVVTKK 295  
 E+YSEG R VV+ K  
 55 Sbjct: 189 ETYSEGQIGRHVVIAPK 206

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 841> which encodes the amino acid sequence <SEQ ID 842>. Analysis of this protein sequence reveals the following:

Possible site: 27  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 55 bacterial cytoplasm --- Certainty=0.3721(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

-330-

Identities = 176/302 (58%), Positives = 223/302 (73%), Gaps = 32/302 (10%)

Query: 23 MVLFTGATVEEAIEKGLQELNISRLRAHIKVVVSREKKGFLGFGKKPAKVEIEGITDEVTD 82  
 MVLFTG TVEEAIE GLQEL +SRL+AHIKV+S+EKKGFLGFGKKPA+V+IEGI+D+  
 5 Sbjct: 1 MVLFTGKTVEEAIEITGLQELGLSRLKAHIKVISKEKKGFLGFGKKPAQVDIEGISDKTVY 60

Query: 83 INESVALKNI-----KNVPS--SVDVVEEYIEEVDETLEKEDVSQPELPKIDDK----- 129  
 + A + + +N P+ S DV E I+ + LE ED L D  
 10 Sbjct: 61 KADKKATRGVPEDINRQNTPAVNSADVEPEIKAT-QRLEAEDTKVVPLMSSESPAQTPTS 119

Query: 130 ---NVVTTSEA-----IEKIDL-----LPNIEVAAAQVTKYVENIYEMDLDATI 171  
 VT ++A +E+ ++ +IE AA +V+ YV IYEMD++AT+  
 Sbjct: 120 NLAETVTETKAQQPSIPVEESEVPQDAGNDGFSKDIEKAAQEVSDYVTKIYEMDIEATV 179

Query: 172 ETTTSKRQINLQIETPEAGRIGYHGKVLKSLQLLAQNYLHDRFSKSFVSINVDYVEH 231  
 ET+ ++RQINLQIETPEAGR+IGYHGKVLKSLQLLAQN+LHDR+SK+FSVS+NVHDYVEH  
 15 Sbjct: 180 ETSNNRRQINLQIETPEAGRVIGYHGKVLKSLQLLAQNFLHDRYSKNFVSLNVHDYVEH 239

Query: 232 RTETLIDFSKKIARRVLETNPEYHMDPMSNSERKTVHKTITATIEGVESYSEGNDPNRFV 291  
 RTETLIDF++K+A+RVLE+ + Y MDPMSNSERK VHKT+++IEGV+SYSEGNDPNR+VV  
 20 Sbjct: 240 RTETLIDFTQKVAKRVLSESGQDYTMDPMSNSERKIVHKTIVSSIEGVDSYSEGNDPNRYV 299

Query: 292 VT 293  
 V+  
 25 Sbjct: 300 VS 301

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 265

30 A DNA sequence (GBSx0290) was identified in *S.agalactiae* <SEQ ID 843> which encodes the amino acid sequence <SEQ ID 844>. This protein is predicted to be 60 kd inner-membrane protein (yidC). Analysis of this protein sequence reveals the following:

Possible site: 42

&gt;&gt;&gt; May be a lipoprotein

INTEGRAL	Likelihood = -7.38	Transmembrane	54 - 70 ( 52 - 75)
INTEGRAL	Likelihood = -5.20	Transmembrane	193 - 209 ( 192 - 211)
INTEGRAL	Likelihood = -3.61	Transmembrane	125 - 141 ( 124 - 144)
INTEGRAL	Likelihood = -2.44	Transmembrane	168 - 184 ( 167 - 184)

----- Final Results -----

bacterial membrane --- Certainty=0.3951(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database:

&gt;GP:CAA78595 GB:Z14225 SpoIIJ [Bacillus subtilis]

Identities = 79/243 (32%), Positives = 142/243 (57%), Gaps = 5/243 (2%)

Query: 1 MKKKLKTFTSLILLTGSLLVACG--RGEVSSHSA TLWEQ-IVYAFAKSIQWLS--FNHSIG 55  
 MK+++ ++ LL C + +++ S W++ +VY ++ I +++ + G  
 50 Sbjct: 1 MKRRIGLLLSMVGVMFLLAGCSSVKEPITADSPHFWDKYVVYPLSELITYVAKLTGDNYG 60

Query: 56 LGIILFTLIIRAIMPLYNMQMKSSQKMQEIQPRLKELQKKYPGKDPDNRLKLNDEMOSM 115  
 L IIL T++IR +++PL Q++SS+ MQ +QP +++L++KY KD + KL E ++  
 55 Sbjct: 61 LSIILVTILIRLLILPLMIKQLRSSKAMQALQPEMQKLEKYSSKDQKTQQLQOETMAL 120

Query: 116 YKAEGVNPNYASVPLLIQLPVLWALFQALTRVSFLKVGTFLSLELSQDPYIILPVLAAL 175  
 ++ GVN P A P+LIQ+P+L + A+ R + +FL +L + DPYIILP++A +  
 60 Sbjct: 121 FQKHGVNPLAGCFPILIQPILIGFYHAIMRTQAISEHSFLWFDLGEKDPYIILPIVAGV 180

Query: 176 FTFLSTWLTNKA AVEKNIALTLMTYVMPFIILVTSFNFASGVVLYWTVSNAFQVFQILL 235

-331-

TF+ L ++N + +M ++MP +I+V + NF + + LYW V N F + Q L+  
 Sbjct: 181 ATFVQQKLMMAGNAQQNPQMAMMLWIMPIMIIVFAINFPAALSLYWVVGNLFMIAQTFLI 240

Query: 236 NNP 238

P

Sbjct: 241 KGP 243

A related GBS sequence was identified <SEQ ID 10783> which encodes amino acid sequence <SEQ ID 10784>.

10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 845> which encodes the amino acid sequence <SEQ ID 846>. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> May be a lipoprotein

15 INTEGRAL Likelihood = -6.32 Transmembrane 198 - 214 ( 197 - 220)  
 INTEGRAL Likelihood = -5.52 Transmembrane 59 - 75 ( 57 - 80)  
 INTEGRAL Likelihood = -4.25 Transmembrane 130 - 146 ( 129 - 150)  
 INTEGRAL Likelihood = -2.28 Transmembrane 173 - 189 ( 170 - 189)

20 ----- Final Results -----

bacterial membrane --- Certainty=0.3527(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:

>GP:BAA05234 GB:D26185 stage III sporulation [Bacillus subtilis]  
 Identities = 90/249 (36%), Positives = 150/249 (60%), Gaps = 6/249 (2%)

30 Query: 16 IVPLVLLLVACG--RGEVTAQSSSGWDQ-LVYLFARAIQWLS--FDGSIGVGIILFTLTI 70  
 +V + +LL C + +TA S WD+ +VY + I +++ + G+ IIL T+ I  
 Sbjct: 13 MVGVFMLLAGCSSVKEPTTADSPHFDKYVVPYPLSELITYVAKLTGDNVGLSIIIVTILI 72

35 Query: 71 RLMLMPLFNMQIKSSQKMQDIQPELRELQRKYAGKDTQTRMKLAESQALYKKYGVNPYA 130  
 RL+++PL Q++SS+ MQ +QPE+++L+ KY+ KD +T+ KL +E+ AL++K+GVNP A  
 Sbjct: 73 RLLILPLMIKQLRSSKAMQALQPEMQKLKEYSSKDQKTQQLQOETMALFQKHGVNPLA 132

40 Query: 131 SLLPLLIQMPVMIALFQALTRVSFLKTGTFLWVELAQHDHLYLLPVLA AVFTFLSTWLTN 190  
 P+LIQMP++I + A+ R + +FLW +L + D Y+LP++A V TF+ L  
 Sbjct: 133 GCFPILIQMPILIGFYHAIMRTQAISEHSFLWFDLGEKDPYYILPIVAGVATFVQQKLMM 192

45 Query: 191 LAAKEKNVMMTVMIIYVMPMLIFFMGFNLAGSVVLYWTVSNAFQVQVQLLLNNP-FKIIAE 249  
 ++N M +M+++MP+MI N + + LYW V N F + Q L+ P K E  
 Sbjct: 193 AGNAQQNPQMAMMLWIMPIMIIVFAINFPAALSLYWVVGNLFMIAQTFLIKGPDIKKNPE 252

45 Query: 250 RQRIANEK 258  
 Q+ ++K  
 Sbjct: 253 PQKAGGKKK 261

An alignment of the GAS and GBS proteins is shown below:

50 Identities = 172/270 (63%), Positives = 217/270 (79%), Gaps = 1/270 (0%)

Query: 1 MKKKLKTFSILLTGSLLVACGRGEVSSHSATLWEQIVYAFAKSIQWLSFNHSGILGII 60  
 +KK +K ++ L LLVACGRGEV++ S++ W+Q+VY FA++IQWLSF+ SIG+GIIL  
 Sbjct: 7 VKKNIKIARIVPLV-LLLVACGRGEVTAQSSSGWDQLVYLFARAIQWLSFSGSIGVGIIL 65

55 Query: 61 FTLLIRAIMMPLYNMQKSSQKMQEIQPRKELQKKYPGKDPDNRLKLNDEMOSMYKAEG 120  
 FTL IR ++MPL+NMQ+KSSQKMQ+IQP L+ELQ+KY GKD R+KL +E Q++YK G  
 Sbjct: 66 FTLLIRLMLMPLFNMQIKSSQKMQDIQPELRELQRKYAGKDTQTRMKLAESQALYKKYG 125

60 Query: 121 VNPYASVLPILLIQLPVLWALFQALTRVSFLKVGTFLSLELSQDPYYILPVLAALFTFLS 180  
 VNPYAS+LPLLIQ+PV+ ALFQALTRVSFLK GTFL +EL+Q D Y+LPVLA+FTFLS  
 Sbjct: 126 VNPYASLLPILLIQMPVMIALFQALTRVSFLKTGTFLWVELAQHDHLYLLPVLA AVFTFLS 185

-332-

Query: 181 TWLTNKA AVEKNIALTLMTYVMPFIILVTSFNFA SGVVLVYWTVSNAFQVFQILLNNPYK 240  
 TWLTN AA EKN+ +T+M YVMP +I FN ASGVVLVYWTVSNAFQV Q+LLLNNP+K  
 Sbjct: 186 TWLTNLA AKEKNVMTVMVIYVMPLMIFFMGEFNLA SGVVLVYWTVSNAFQVQVQLLLNNPFK 245

Query: 241 IIKVREEAVRVAHEKEQQRVKRAKRKASKKR 270  
 II R+ E+ R +RA++KA K++  
 Sbjct: 246 IIAERQRLANEKEKERRLRERRARKKAMKRK 275

10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8527> and protein <SEQ ID 8528> were also identified. Analysis of this protein sequence reveals the following:

15 Lipop: Possible site: 20 Crend: 5  
 McG: Discrim Score: 4.90  
 GvH: Signal Score (-7.5): -0.39  
 Possible site: 42  
 >>> May be a lipoprotein  
 20 ALOM program count: 4 value: -7.38 threshold: 0.0  
 INTEGRAL Likelihood = -7.38 Transmembrane 54 - 70 ( 52 - 75)  
 INTEGRAL Likelihood = -5.20 Transmembrane 193 - 209 ( 192 - 211)  
 INTEGRAL Likelihood = -3.61 Transmembrane 125 - 141 ( 124 - 144)  
 INTEGRAL Likelihood = -2.44 Transmembrane 168 - 184 ( 167 - 184)  
 25 PERIPHERAL Likelihood = 2.54 217  
 modified ALOM score: 1.98

\*\*\* Reasoning Step: 3

----- Final Results -----

30 bacterial membrane --- Certainty=0.3951(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

35 32.8/62.3% over 242aa

Bacillus subtilis

EGAD|17722| stage III sporulation protein j precursor Insert characterized

OMNI|NT01BS4782 -identity Insert characterized

40 SP|Q01625|SP3J\_BACSU STAGE III SPORULATION PROTEIN J PRECURSOR. Edit characterized

GP|40023|emb|CAA44401.1||X62539 unnamed protein product Insert characterized

GP|467388|dbj|BAA05234.1||D26185 stage III sporulation Insert characterized

GP|2636651|emb|CAB16141.1||Z99124 alternate gene name: spo0J87 Insert characterized

PIR|I40437|I40437 stage III sporulation protein spoIIJ - Insert characterized

45 ORF02221(301 - 1014 of 1413)

EGAD|17722| S4098(3 - 245 of 261) stage III sporulation protein j precursor { acillus subtilis|OMNI|NT01 S4782 -identitySP|Q01625|SP3J\_ ACSU STAGE III SPORULATION PROTEIN J PRECURSOR.GP|40023|emb|CAA44401.1||X62539 unnamed protein product { acillus subtilis|GP|467388|dbj| AA05234.1||D26185 stage III sporulation { acillus subtilis|GP|2636651|emb|CA 16141.1||Z99124 alternate gene name: spo0J87 { acillus subtilis|PIR|I40437|I40437 stage III sporulation protein spoIIJ - acillus subtilis

50 %Match = 17.0  
 %Identity = 32.8 %Similarity = 62.2  
 Matches = 79 Mismatches = 88 Conservative Sub.s = 71

55 219 249 279 309 339 393 420  
 DFVVIARKGVEELDYQALEKNLIHVLKIAGLI\*KGIKLKKLKTFSLLLTGSLLVACG--RGEVSSH SATLWEQ-IVYA  
 : ||::: : : : || : : : : : : : : : :  
 MLLKRRIGLLSMVGVFLLAGCSSVKEPITADSPHFWDKYVVP  
 60 10 20 30 40

474 504 534 564 594 624 654  
 FAKSIQWLS--FNHSIGLGIILFTLIIRAIMMPLYNMQMKSSQKMQEIQPRLKELOKKYPGKDPDNRLKLNDQMOSMYKA

-333-

```

::: | ::: : || ||| |::|| :::|| |::||: || :|| ::|::|| || : || | ::::
LSELITYVAKLTGDNVGLSIIIVTILIRLLILPLMIKQLRSSKAMQALQPEMQKLKEKYSSKDQKTQQKLQOETMALFQK
60 70 80 90 100 110 120

5 684 714 744 774 804 834 864 894
EGVNPYASVLPPLLIQLPVLWALFQALTRVSLKVGTFSLSELSQPDPPYIILPVLAALFTFLSTWLTNKAAVEKNIALTLM
||| | :|:|:|:|:| ::|:| : : :| :| : |||||:|:| : ||: | :| : :|
HGVNPLAGCFPILIQMPILIGFYHAIMRTQAISEHSFLWFDLGEKDPYIILPIVAGVATFVQQLMMAGNAQQNPQMAMM
140 150 160 170 180 190 200

10 924 954 984 1014 1044 1074 1104 1134
TYVMPFIILVTSFNFAAGVVLVTVSNFQVQFQILLNNPYKIIKVREEAVRVAHEKEQVRKRAKRKASKKRK*ENHGII
::|| :|:| : || : : ||| | | :|:| : : :
LWIMPIMIIVFAINFPAALSLYVVVGNLFMIAQTFLIKGPDIKKNPEPQKAGGKKK
15 220 230 240 250 260

```

**Example 266**

A DNA sequence (GBSx0291) was identified in *S.agalactiae* <SEQ ID 847> which encodes the amino acid sequence <SEQ ID 848>. Analysis of this protein sequence reveals the following:

```

Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3778(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9579> which encodes amino acid sequence <SEQ ID 9580> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CAA44400 GB:X62539 homologous to E.coli rnpA [Bacillus subtilis]
Identities = 52/109 (47%), Positives = 77/109 (69%), Gaps = 1/109 (0%)

Query: 21 LKKTyrVKSDKDFQMFISRGKNVANRKFVIYYLEK-EQKHFRVGVISVSKKLGNAVVRNAI 79
LKK R+K ++DFQ +F G +VANR+FV+Y L++ E RVG+SVSKK+GNAV+RN I
Sbjct: 4 LKKRNRLKKNEDFQKVFKHGTSVANRQFVLYTLDPENDELRVGLSVSKKIGNAVMRNRI 63

Query: 80 KKRIRHVLLSQKTALQDYDFVVIARKGVEELDYQALEKNLIHVLKIAGL 128
KR IR L +K L++ D+++IARK +L Y+ +K+L H+ + + L
Sbjct: 64 KRLIRQAFLEEKERLKEKDYIIIRKPKASQLTYEETKKSLOHLFRKSSL 112

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 849> which encodes the amino acid sequence <SEQ ID 850>. Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3820(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 73/109 (66%), Positives = 88/109 (79%)

Query: 21 LKKTyrVKSDKDFQMFISRGKNVANRKFVIYYLEKEQKHFRVGVISVSKKLGNAVVRNAIK 80
LKKTyrvK +KDFQ IF GK+ ANRKFVIY+L + Q HFRVGVISV KK+GNAV RNA+K
Sbjct: 1 LKKTyrvKREKDFQAIFKDGKSTANRKFVIYHLNRQDHFVRVGVISVGGKIGNAVTRNAVK 60

Query: 81 RKIRHVLLSQKTALQDYDFVVIARKGVEELDYQALEKNLIHVLKIAGLI 129
RKIRHV+++ L+ DFVVIARKGV L+YQ L++NL HVLK+A L+

```

-334-

Sbjct: 61 RKIRHVIMALGHQLKSEDFVVIARKGVHSLEYQELQQNLHHVLKLAQLL 109

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 5 Example 267

A DNA sequence (GBSx0292) was identified in *S.agalactiae* <SEQ ID 851> which encodes the amino acid sequence <SEQ ID 852>. This protein is predicted to be glycerol-3-phosphate dehydrogenase, NAD-dependent (gpsA). Analysis of this protein sequence reveals the following:

10 Possible site: 33  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1429(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 15 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8529> which encodes amino acid sequence <SEQ ID 8530> was also identified. There is a signal peptide at residues 1-19. The protein has homology with the following sequences in the GENPEPT database:

20 >GP:AAA86746 GB:U32164 NAD(P)H-dependent dihydroxyacetone-phosphate  
 reductase [Bacillus subtilis]  
 Identities = 177/333 (53%), Positives = 241/333 (72%)

25 Query: 18 QKIAVLGPGSWGTTALAQVLNDNGHEVRLWGNVVEQIEEINTNHTNQRYFKDITLDSKIK 77  
 +K+ +LG GSWGTTALA VL DNG+EV +W + + I +IN H N+ Y ++ L + IK  
 Sbjct: 2 KKVITMLGAGSWGTTALALVLTNDNGNEVCVWAHRADLIHQINELHENKDYLPNVKLSTSIK 61

30 Query: 78 YTNLEEAINNVDSILFVVPTKVTTRLVAKQVANLLKHKVVLMHASKGLEPGTHERLSTILE 137  
 T+++EA+++ D I+ VPTK R V +Q + K V +H SKG+EP + R+S I+E  
 Sbjct: 62 TTDMKEAVSDADVIIIVAVPTKAIREVLRQAVPFITTKAVFVHVS KGIEPDSLRLRISEIME 121

35 Query: 138 EEISEQYRSDIVVVGSPSHAEAAIVRDITLITAASKDIEAAKYVQKLFNSHYFRLYTNTD 197  
 E+ R DIVV+SGPSHAE +R T +TA+SK + AA+ VQ LF NH FR+YTN D  
 Sbjct: 122 IELPSDVRRDIVVLSGSPSHAEVGLRHATTVTASSKSMRAAEVQDLFINHNFRVYTNP 181

40 Query: 198 VVGVTAGALKNIIAIVGAGALHGLGYGDNAKAAIITRGLAEITRLGVQLGADPLTFSGLS 257  
 ++GVE GALKNIIA+ AG GLGYGDNAKAA+ITRGLAEI RLG ++G +PLTFSGL+  
 Sbjct: 182 IIGVEIGGALKNIIAALAGITDGLGYGDNAKAAITRGLAEIARLGTMGGNPLTFSGLT 241

45 Query: 258 GVGD LIVTTSVHSRNRWAGDALGRGEKLEDIEKNMGMVIEGISTTKVAYEIAQNINVM 317  
 GVGD LIVT TSVHSRNRWAG+ LG+G KLED+ + MGMV+EG+ TTK AY++++ +V M  
 Sbjct: 242 GVGD LIVTCTSVHSRNRWAGNLLGKGYKLEDVLEEMGMVVEGVRTTKAAYQLSKKYDV 301

Query: 318 PITEAIYKSIYEGANIKDSILDMSNEFRSENE 350  
 PITEA+++ ++ G ++ ++ +M+ E E  
 Sbjct: 302 PITEALHQVLFNGQKVETAVESLMARGKTHEME 334

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 853> which encodes the amino acid sequence <SEQ ID 854>. Analysis of this protein sequence reveals the following:

50 Possible site: 19  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.0882(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 55 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:



-335-

Identities = 287/338 (84%), Positives = 316/338 (92%)

Query: 15 MTKQKIAVLGPGSWGTTALAQVLNDNGHEVRLWGNVVEQIEEINTNHTNQRYFKDITLDSK 74  
 Sbjct: 1 MTKQK+A+LGPGSWGTTAL+QVLNDNGH+VRLWGN+ +QIEEINT HTN+ YFKDI LD

Query: 75 IKAYTNLEEAINNVDSILFVVPTKVTRLVAKQVANLLKHKVVLMHASKGLEPGTHERLST 134  
 I A +L +A+++VD++LFVVPTKVTRLVA+QVA +L HKVV+MHASKGLEP THERLST  
 Sbjct: 61 ITATLDLGQALSDVDAVLVFPVPTKVTRLVARQVAAILDHKVVVMHASKGLEPETHERLST 120

Query: 135 ILEEEISEQYRSDIVVVSGPSHAEEAIVRDITLITAASKDIEAAKYVQKLFNSHYFRLYT 194  
 ILEEEI +RS++VVVSGPSHAEE IVRDITLITAASKDIEAAKYVQ LFSNHYFRLYT  
 Sbjct: 121 ILEEEIPAHRSEVNVVSGPSHAETIVRDITLITAASKDIEAAKYVQSLFNSHYFRLYT 180

Query: 195 NTDVVGVEVETAGALKNIIAVGAGALHGLGYGDNAKAAIITRGLAEITRLGVQLGADPLTFS 254  
 NTDV+GVETAGALKNIIAVGAGALHGLGYGDNAKAA+ITRGLAEITRLGV+LGADPLT+S  
 Sbjct: 181 NTDVIGVETAGALKNIIAVGAGALHGLGYGDNAKAAVITRGLAEITRLGVKLADPLTYS 240

Query: 255 GLSGVGDLIVTGTSVHSRNRWAGDALGRGEKLEDIEKNMGMVIEGISTTKVAYEIAQNLN 314  
 GLSGVGDLIVTGTSVHSRNRWAG ALGRGEKLEDIE+NMGMVIEGI+TTKVAYEIAQ+L  
 Sbjct: 241 GLSGVGDLIVTGTSVHSRNRWAGALGRGEKLEDIERNMGMVIEGIATTKVAYEIAQDLG 300

Query: 315 VYMPITEAIYKSIYEGANIKDSILDMSNEFRSENEWH 352  
 VYMPIT AIYKSIYEGA+IK+SIL MMSNEFRSENEWH  
 Sbjct: 301 VYMPITTAIYKSIYEGADIKESILGMSNEFRSENEWH 338

SEQ ID 8530 (GBS291) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 59 (lane 5; MW 38.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 2; MW 64kDa).

GBS291-GST was purified as shown in Figure 226, lane 10-11.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 268

A DNA sequence (GBSx0293) was identified in *S.agalactiae* <SEQ ID 855> which encodes the amino acid sequence <SEQ ID 856>. This protein is predicted to be glucose-1-phosphate uridylyltransferase (gtaB). Analysis of this protein sequence reveals the following:

Possible site: 25  
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA28714 GB:AB001562 glucose-1-phosphate uridylyltransferase  
 [Streptococcus mutans]  
 Identities = 263/296 (88%), Positives = 285/296 (95%)

Query: 2 KVRKAVIPAAGLGTRFLPATKALAKEMLPIDVKPTIQFIVEEALKSGIEDILVVTGKSKR 61  
 Sbjct: 5 KVRKAVIPAAGLGTRFLPATKALAKEMLPIDVKPTIQFIVEEALKSGIEDILVVTGKSKR 64

Query: 62 SIEDHFDNSNFELEYNLKEKGKNELLKLVDETTGLRHLFIRQSHPRGLGDAVLQAKAFVGN 121  
 SIEDHFDNSNFELEYNL++KGK +LLKLV++TT I LHFIRQSHPRGLGDAVLQAKAFVGN  
 Sbjct: 65 SIEDHFDNSNFELEYNLEQKGKTDLLKLVDTTAINLHFIHQSHPRGLGDAVLQAKAFVGN 124

Query: 122 EPFVVMGLGDDLMIDITNNKVIPLTKQLINDFEATHASTIAMEVPHEDVSAYGVIAPQGE 181

-336-

EPFVVMGLGDDLMDIT++K IPLT+QL+ND+E THASTIAVMEVPHEVDVSAYGVIAPQGEG  
 Sbjct: 125 EPFVVMGLGDDLMDITDDKAIPLTRQLMNDYEETHASTIAVMEVPHEVDVSAYGVIAPQGEG 184

Query: 182 VNGLYSVNTFVEKPSPEEAPSNLAIIGRYLLTPEIFNILETQKPGAGNEIQLTDAIDTLN 241  
 V+GLYSV+TFVEKP+P+EAPSNLAIIGRYLLTPEIF ILETQ+PGAGNE+QLTDAIDTLN  
 Sbjct: 185 VSGLYSVDTFVEKPAPKEAPSNLAIIGRYLLTPEIFTILETQEPGAGNEVQLTDAIDTLN 244

Query: 242 KTQRVFARKFTGDRYDVGDKFGFMKTSIDYALQHPQVKDDLKYYIIDLGKSLEKTS 297  
 KTQRVRFAR+F G RYDVGDKFGFMKTSIDYAL+HPQVK+DLK YII+LGK L++ S  
 Sbjct: 245 KTQRVFAREFKGRYDVGDKFGFMKTSIDYALKHPQVKEDLKAYIIBLGKKLDQKS 300

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 857> which encodes the amino acid sequence <SEQ ID 858>. Analysis of this protein sequence reveals the following:

Possible site: 26  
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 257/295 (87%), Positives = 277/295 (93%)

Query: 2 KVRKAVIPAAGLGTRFLPATKALAKEMLPIDVKPTIQFIVEEALKSGIEDILVVTGKSKR 61  
 KVRKA+IPAAGLGTRFLPATKALAKEMLPIDVKPTIQFIVEEALKSGIE+ILVVTGK+KR  
 Sbjct: 3 KVRKAIIPAAGLGTRFLPATKALAKEMLPIDVKPTIQFIVEEALKSGIEEILVVTGAKR 62

Query: 62 SIEDHFDNSNFELEYNLKEKGNELLKLVDETT GIRLHFIRQSHPRGLGDAVLQAKAFVGN 121  
 SIEDHFDNSNFELEYNL+ KGNELLKLVDETT I LHFIRQSHPRGLGDAVLQAKAFVGN  
 Sbjct: 63 SIEDHFDNSNFELEYNLQAKGNELLKLVDETTAINLHFIRQSHPRGLGDAVLQAKAFVGN 122

Query: 122 EPFVVMGLGDDLMDITNNKVIPLTKQLINDFEATHASTIAVMEVPHEVDVSAYGVIAPQGEG 181  
 EPFVVMGLGDDLMDITN PLTKQL+ D++ THASTIAVM+VPHEVDVS+YGVIAPOG+  
 Sbjct: 123 EPFVVMGLGDDLMDITNASAKPLTKQLMEDYDKTHASTIAVMKVPHEVDVSSYGVIAPOGKA 182

Query: 182 VNGLYSVNTFVEKPSPEEAPSNLAIIGRYLLTPEIFNILETQKPGAGNEIQLTDAIDTLN 241  
 V GLYSV+TFVEKP PE+APS+LAIIGRYLLTPEIF ILE Q PGAGNE+QLTDAIDTLN  
 Sbjct: 183 VKGLYSVDTFVEKPPEDAPSDLAIIGRYLLTPEIFGILERQTPGAGNEVQLTDAIDTLN 242

Query: 242 KTQRVFARKFTGDRYDVGDKFGFMKTSIDYALQHPQVKDDLKYYIIDLGKSLEKT 296  
 KTQRVRFAR+F G+RYDVGDKFGFMKTSIDYAL+HPQVK+DLK YII LGK+LEK+  
 Sbjct: 243 KTQRVFAREFKGNRYDVGDKFGFMKTSIDYALEHPQVKEDLKNYIIKLGKALEKS 297

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 269

A DNA sequence (GBSx0294) was identified in *S.agalactiae* <SEQ ID 859> which encodes the amino acid sequence <SEQ ID 860>. Analysis of this protein sequence reveals the following:

Possible site: 42  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -4.94 Transmembrane 28 - 44 ( 27 - 45)

----- Final Results -----

bacterial membrane --- Certainty=0.2975(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

-337-

>GP:CAB15143 GB:Z99120 similar to ABC transporter (lipoprotein)  
[Bacillus subtilis]

Identities = 148/346 (42%), Positives = 222/346 (63%), Gaps = 16/346 (4%)

5 Query: 31 LTLTSLSVLTTLTACGNRSDKSAN---KSDIKVAMVTNQGGVDDKSFNQSAWEGLOKWK 87  
++L+ + L ACGN S + K+ VAMVT+ GGVDDKSFNQSAWEG+Q +GK+  
Sbjct: 1 MSLVIAAGTILGACGNSEKSSGSGEGKNKFSVAMVTDVGGVDDKSFNQSAWEGIQAFGKE 60

10 Query: 88 KGLTKG-NGFDYFQSSNESDHANNLDTAASSGYNLIFGIGFGLHDTIEKVSENNKDVKYV 146  
GL KG NG+DY QS +++D+ NL+ A ++LI+G+G+ + D+I ++++ K+ +  
Sbjct: 61 NGLKKGKNGYDYLQSKSDADYTTNLNKLARENFDLIYGVGYLMEDSISEIADQRKNTNFA 120

15 Query: 147 IVDDIIKGENVASVTFADNEAAYLAGVAAAKTTKTQVGFVGGMEGVVVKRFEAGFKAG 206  
I+D ++ K+NVAS+TF + E ++L GVAAA ++K+ +GF+GGME ++K+FE GF+AG  
Sbjct: 121 IIDAVVD-KDNVASITFKEQEGSFLVGVAALSSKSGKIGFVGGMESELIKKEVGFVFRAG 179

20 Query: 207 VKSIDPAIKVAVSYAGSFTDAAGKKTIAATQYATGVDVIYQAAGGTGAGIFSEAKTENET 266  
V++++P V V YAG F A GK A + Y +GVDVIY +AG TG G+F+EAK  
Sbjct: 180 VQAVNPKA VVEVKYAGGFADKADVGKATAESMYKSGVDVIYHSAGATGTGVFTEAK---NL 236

25 Query: 267 RKESNK--VWVIGVDRDQSQEGNYVSKDGKKANFVLASTIKEVGKSLQSVAELETKKQYP 324  
+KE K VWVIGVD+DQ EG +G N L S +K+V ++ V + ++P  
Sbjct: 237 KKEDPKRDVWVIGVDKDYAEGQV---EGTDDNVTLTSMVKKVDTVVEDVTKKASDGKFP 293

30 Query: 325 GGVTVFGLKDSGVDI--KEHQLSSEGSVAVKKAKEDIVSGKIQVP 368  
GG+ +GL GV I + LS + AV K K+ I+ G +++P  
Sbjct: 294 GGETLTYGLDQDGVGISPSKQNLSDDVIAVDKWKKKIIDG-LEIP 338

There is also homology to SEQ ID 862.

30 A related GBS gene <SEQ ID 8531> and protein <SEQ ID 8532> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: 20 Crend: 3  
Sequence Pattern: CGNR  
SRCFLG: 0

35 McG: Length of UR: 19  
Peak Value of UR: 2.31  
Net Charge of CR: 2  
McG: Discrim Score: 5.09  
GvH: Signal Score (-7.5): -3.29

40 Possible site: 19  
>>> May be a lipoprotein  
Amino Acid Composition: calculated from 21  
ALOM program count: 0 value: 5.20 threshold: 0.0  
PERIPHERAL Likelihood = 5.20 90

45 modified ALOM score: -1.54

\*\*\* Reasoning Step: 3

50 ----- Final Results -----  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

55 52.8/73.9% over 239aa

Listeria  
monocytogenes  
SP|Q48754| CD4+ T CELL-STIMULATING ANTIGEN PRECURSOR. Insert characterized  
GP|7240601|gb|AAB35725.2||S80336 CD4+ T cell-stimulating antigen Insert characterized

60 ORF02225(385 - 1086 of 1710)  
SP|Q48754|TCSA\_LISMO(8 - 247 of 268) CD4+ T CELL-STIMULATING ANTIGEN  
PRECURSOR.GP|7240601|gb|AAB35725.2||S80336 CD4+ T cell-stimulating antigen {Listeria  
monocytogenes}

-338-

%Match = 21.7  
 %Identity = 52.7 %Similarity = 73.8  
 Matches = 125 Mismatches = 59 Conservative Sub.s = 50

```

5      294      324      354      384      414      444      465      489
NFLWEK*NKVC*MIFLCYDRNLFLCDYNLLGGSFVSNRKIIGLTLLSLSVLTLTACGNRSD--KSANKS--DIKVAMVT
      : |::: | : | ||: ||      | :|| | ||||
      MKKRTFALALSMIIASGVILGACGSSDDKKSSDDKSSKDFTVAMVT
      10      20      30      40

10     519      549      579      606      636      666      696      726
NQGGVDDKSFNQSAWEGLOKWKKKGLTKG-NGFDYFQSSNESDHANNLDTAASSGYNLIFGIGFGLHDTIEKVSENNKD
      : |||||:|||||:|||||:| : || :||:|:|:|:|:|:| : ||:|:|:|:|:| : ||:|:|:
15     DTGGVDDRSFNQSAWEGLOKFGKANDMEKGTGDGYNYLQSASEADYKTNLNTAVRSDYDLTYGIGYKLKDAIEEVSKQKPK
      60      70      80      90      100      110      120

20     756      786      816      846      876      906      936      966
VKYVIVDDIIKGKENVASVTFADNEAAYLAGVAAAKTTKTKTVGFIGGMEGVVVKRFEAGFKAGVKSIDPAIKVASVYAG
      :: |||| | :|| | : ||: || | |||| |||:|:|:| : ||||| ||||:| : : ||
25     NQFAIVDDTIDDRDNVVSIGFKDNDGSYLVGVVAGLTTKTNKVGVGVGKGTVIDRFEAGFTAGVKAVNPNAQIDVQYAN
      140      150      160      170      180      190      200

996      1026      1056      1086      1116      1146      1176      1206
SFTDAAGKGTIAATQYATGVDVIYQAAGGTGAGIFSEAKTENETRKESNKKVWVIGVDRDQSQEGNYVSKDGKANFVLAS
      | | ||: ||:| :|||:|:|:|:|:|:| :|:|:| : : :
25     DPAKADKGQIASSMYSSGVDVIFHAAGGTGNGVFAEAKNLKKKDLQMPVPGNSKLGCFGG
      220      230      240      250      260

```

A related GBS nucleic acid sequence <SEQ ID 10947> which encodes amino acid sequence <SEQ ID 10948> was also identified.

SEQ ID 8532 (GBS108) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 7; MW 39.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 9; MW 64.6kDa).

The GBS108-GST fusion product was purified (Figure 202, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 273), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 270

A DNA sequence (GBSx0295) was identified in *S.agalactiae* <SEQ ID 863> which encodes the amino acid sequence <SEQ ID 864>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
45     INTEGRAL    Likelihood = -12.74    Transmembrane  206 - 222 ( 197 - 224)
      INTEGRAL    Likelihood = -3.72     Transmembrane  174 - 190 ( 171 - 194)
      INTEGRAL    Likelihood = -3.19     Transmembrane   98 - 114 (  98 - 116)
      INTEGRAL    Likelihood = -1.54     Transmembrane  120 - 136 ( 120 - 139)
      INTEGRAL    Likelihood = -0.90     Transmembrane  157 - 173 ( 157 - 173)

50     ----- Final Results -----
      bacterial membrane --- Certainty=0.6095(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

-339-

>GP:CAB90755 GB:AJ400707 hypothetical protein [Streptococcus uberis]  
Identities = 126/218 (57%), Positives = 166/218 (75%)

5 Query: 8 KEYPTTVLLVSLTTLVFLLMQLTYGSAESSQVIFQFGGIQGDYLKAYPTNLWRLISPIF 67  
KE P T +S+T L+F++MQ+ YGS A+S QV+FQFGG+ G +K+ P+ LWRL++PIF  
Sbjct: 5 KEKPVTFFFLSVTILLFIVMQVYFGSWAKSPQVVFQFGGMFGLVVKSMPSQLWRLVTPIF 64

10 Query: 68 VHIGWEHFLNLGLALYFVGQMGESIWSLRFLLILYLSGLMGNIFTLFFTPHVVAAGAST 127  
+HIGWEHFL+N L LYFVGQ+ ESIWGS FL+LY+LSG+MGN+ TLFFTPHVVAAGAST  
Sbjct: 65 IHIGWEHFLINSITLYFVGQLAESIWSRFFLLLYVLSGIMGNVLTLLFFTPHVVAAGAST 124

15 Query: 128 SLFGVFSIAIAIAGYFGKNPYLKQVGKSYQVMILLNFFNIPTPGVSLAGHVGGVLV 187  
SLFG+F+AI + GYFG N LK +GKSYQ +I+LNL N+F P V + GH+GG +GG L  
Sbjct: 125 SLFGLFAAIVVVGYPFGHNQLLSIGKSYQTLIILNLMNLFMPNVGIVGHLGGALGGALA 184

Query: 188 AIFLTKQNGSLLFKTWQSILALMIFIIVSISLIGLSLV 225  
A+FL + LF Q AL+ ++ +++ LI LSL+  
Sbjct: 185 AVFLPTLLDAELFTKKQKTSALLSYLTALVLITLSLM 222

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 865> which encodes the amino acid sequence <SEQ ID 866>. Analysis of this protein sequence reveals the following:

Possible site: 43  
>>> Seems to have no N-terminal signal sequence

25 INTEGRAL Likelihood = -9.92 Transmembrane 214 - 230 ( 212 - 232)  
INTEGRAL Likelihood = -5.36 Transmembrane 135 - 151 ( 128 - 153)  
INTEGRAL Likelihood = -1.81 Transmembrane 101 - 117 ( 100 - 117)  
INTEGRAL Likelihood = -1.44 Transmembrane 183 - 199 ( 182 - 199)  
INTEGRAL Likelihood = -0.53 Transmembrane 166 - 182 ( 166 - 182)

30 ----- Final Results -----  
bacterial membrane --- Certainty=0.4970(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the databases:

>GP:CAB90755 GB:AJ400707 hypothetical protein [Streptococcus uberis]  
Identities = 72/128 (56%), Positives = 94/128 (73%)

40 Query: 106 FLLLYVLSGVMGNAFTFWLTTPETVAAGASTSLFGLFAAIVVLSFLGKNQALKDLGKSYQT 165  
FLLLYVLSG+MGN T + TP VAAGASTSLFGLFAAIVV+ + G NQ LK +GKSYQT  
Sbjct: 95 FLLLYVLSGIMGNVLTLLFFTPHVVAAGASTSLFGLFAAIVVVGYPFGHNQLLSIGKSYQT 154

45 Query: 166 LIVVNLLMNLMPNVSMAGHIGGVVGALLSIVFPPTKMRVITVKKTKRMLALVSYGIILV 225  
LI++NL+MNLMPNV + GH+GG +GGAL ++ PT + K ++ AL+SY + +  
Sbjct: 155 LIILNLMNLFMPNVGIVGHLGGALGGALAAVFLPTLLDAELFTKKQKTSALLSYLTAL 214

Query: 226 GVLVLGFL 233  
++ L +  
Sbjct: 215 VLITLSLM 222

50

An alignment of the GAS and GBS proteins is shown below:

Identities = 63/132 (47%), Positives = 92/132 (68%)

55 Query: 94 GSLRFLILYLSGLMGNIFTLFFTPHVVAAGASTSLFGVFSIAIAIAGYFGKNPYLKQVGK 153  
G FL+LY+LSG+MGN FT + TP VAAGASTSLFG+F+AI + + GKN LK +GK  
Sbjct: 102 GLTPFLLLYVLSGVMGNAFTFWLTTPETVAAGASTSLFGLFAAIVVLSFLGKNQALKDLGK 161

60 Query: 154 SYQVMILLNLFNIFTPTPGVSLAGHVGGVLVVAIFLTKQNGSLLFKTWQSILALMIFI 213  
SYQ +I++NL N+F P VS+AGH+GG+VGG L++I + + K + +LAL+ +  
Sbjct: 162 SYQTLIVVNLLMNLMPNVSMAGHIGGVVGALLSIVFPPTKMRVITVKKTKRMLALVSYG 221

Query: 214 IVSISLIGLSLV 225  
I+ + ++ L +  
Sbjct: 222 IILVGVLVLGFL 233

-340-

A further corresponding DNA sequence was identified in *S.pyogenes* <SEQ ID 9083> which encodes the amino acid sequence <SEQ ID 9084>. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -7.70 Transmembrane 12 - 28 ( 7 - 30)

----- Final Results -----

bacterial membrane --- Certainty=0.4079(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS sequences follows:

Score = 74.5 bits (180), Expect = 5e-16  
Identities = 37/96 (38%), Positives = 48/96 (49%)

Query: 1 MTQLLKRYPXXXXXXXXXXXXXXXXXAMQVYGHLLATGAQAIYQVGGMFGLLVKAMPDQLWRL 60  
M + K YP MQ+ YG A +Q I+Q GG+ G +KA P LWRL

Sbjct: 3 MKKFAKEYPTTLLVSLTTLVFLMLQTYGSQAESSQVIFQFGGIQGDYLLKAYPTNLWRL 62

Query: 61 VTPXXXXXXXXXXVNGLTLYFVGQIVEDLWGSRLF 96  
++P +NGL LYFVGQ+ E +WGS F

Sbjct: 63 ISPIFVHIGWEHFLNGLALYFVGQMGESIWGSRLF 98

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 271

A DNA sequence (GBSx0296) was identified in *S.agalactiae* <SEQ ID 867> which encodes the amino acid sequence <SEQ ID 868>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2055(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA28715 GB:AB001562 hypothetical protein [Streptococcus mutans]  
Identities = 96/173 (55%), Positives = 129/173 (74%)

Query: 1 MEKLLRKEVLITLKSQPQAYKSEVDCKLLEAFIKTKAYQNSCVIATYLSFDYEYNTQLL 60  
M KK R +V+ LK Q +A K D +LLE I+ +AYQ + VIATYL+F +E++T LL

Sbjct: 1 MMKKDYRTQVIEDLKKQDKAKKVLRLDEQLLEELIQLEAYQKAHVIATYLAFFPEFDTSL 60

Query: 61 IKQALCDGKRVLPKTYPKGKMI FVDYQKDNLRTPFGLLEPVNDRAVEKASIDLIHVP 120  
I+QA D K ++VPKTYP+ KMLFV Y + +L+ T FGL EP ++ A+EK++IDLIHVP

Sbjct: 61 IEQAQRDNKSIVVPKTYPQRKMI FVVYDEADLQITKFGLEPRSEEALEKSAIDLIHVP 120

Query: 121 LIFNNKGRIGYGAGYFDRYLSDFEGDTISTYRCQRQDFVEEKHDAVAVKEVL 173  
L FNN+G+RIG+GAGY+D+YL+DF+GDT+STIY Q+ F D+ VKEVL

Sbjct: 121 LAFNNEGYRIGFGAGYYDQYLADFQGDIVSTIYSFQQFTFEPSSFDDIPVKEVL 173

A related GBS nucleic acid sequence <SEQ ID 10925> which encodes amino acid sequence <SEQ ID 10926> was also identified.

-341-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 272

- 5 A DNA sequence (GBSx0297) was identified in *S.agalactiae* <SEQ ID 869> which encodes the amino acid sequence <SEQ ID 870>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

10 INTEGRAL Likelihood = -1.44 Transmembrane 161 - 177 ( 161 - 177)  
INTEGRAL Likelihood = -0.22 Transmembrane 29 - 45 ( 28 - 45)

----- Final Results -----

15 bacterial membrane --- Certainty=0.1574(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9305> which encodes amino acid sequence <SEQ ID 9306> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

20 >GP:AAD33517 GB:AF132127 glucose-6-phosphate isomerase  
[Streptococcus mutans]  
Identities = 344/401 (85%), Positives = 374/401 (92%)

25 Query: 1 MDLPENYDKKEFSRIQKAAEKIKSDSEVLVVIGIGGSYLGAKAAIDFLNNHFANLQTAE 60  
++LP+NYDKKEF+RI+KAAEKIKSDSEVLVVIGIGGSYLG+AAIDFLN+ F NL+ EE  
Sbjct: 49 LNLQPNYDKKEFARIKAAEKIKSDSEVLVVIGIGGSYLGARAADFLNSSSFVNLENKEE 108

30 Query: 61 RKAPQILYAGNSISSSTYLADLVEYVQDKEFSVNVISKSGTTTEPAIAFRVFKELLVKKYG 120  
RKAPQILYAGNSISS YLADLV+YV DK+FSVNVISKSGTTTEPAIAFRVFK+LLVKKYG  
Sbjct: 109 RKAPQILYAGNSISSNYLADLVYVADKDFSVNVISKSGTTTEPAIAFRVFKDLLVKKYG 168

35 Query: 121 QEEANKRIYATTDKVGAVKVEADANNWETFVVPDNGGRFSVLTAVGLLPPIAASGADIT 180  
QEEAN+RIYATTD+VKGAVKVEADAN WETFVVPD+VGGRF+VLTAVGLLPPIAASGAD+  
Sbjct: 169 QEEANQRIYATTDKVGAVKVEADANGWETFVVPDNGGRFTVLTAVGLLPPIAASGADLD 228

40 Query: 181 ALMEGANAARKDLSSDKISENIAYQYAAVRNVLYRKGYITEILANYEPSLQYFGEWWKQL 240  
LM GA AAR+D SS ++SEN AYQYAA+RN+LYRKGY+TE+LANYPEPSLQYF EWWKQL  
Sbjct: 229 QLMAGAEARQDYSSAELSENEAYQYAAIRNLYRKGYVTEVLANYEPSLQYFSEWWKQL 288

45 Query: 241 AGESEGDQKQGIYPTSANFSTDHLHSLGQFIQEGYRNLFFETVVRVEKPRKNVTIPELTEDL 300  
AGESEGDQKQGIYPTSANFSTDHLHSLGQFIQEG RNLFFETV+RVEK RKN+ +PE EDL  
Sbjct: 289 AGESEGDQKQGIYPTSANFSTDHLHSLGQFIQEGNRNLFFETVIRVEKARKKNILVPEAAEDL 348

Query: 301 DGLGYLQKGDVDFVNKKATDGVLLAHTDGGVPMFVTLPTQDAYTLGYTIYFFELAIGLS 360  
DGL YLQKGDVDFVNKKATDGVLLAHTDGGVPM F+T+P QD +TLGY IYFFELAIGLS  
Sbjct: 349 DGLAYLQKGDVDFVNKKATDGVLLAHTDGGVPNTFLTPEQDEFTLGYVIYFFELAIGLS 408

50 Query: 361 GYLNSVNPFDQPGVEAYKRNMFALLGKPGFEELS AELNARL 401  
GYLN VNPFDQPGVEAYK+NMFALLGKPGFEEL AELNARL  
Sbjct: 409 GYLNGVNPFDQPGVEAYKKNMFALLGKPGFEELGAELNARL 449

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 871> which encodes the amino acid sequence <SEQ ID 872>. Analysis of this protein sequence reveals the following:

Possible site: 31

55 >>> Seems to have no N-terminal signal sequence  
INTEGRAL Likelihood = -1.44 Transmembrane 209 - 225 ( 209 - 225)

-342-

INTEGRAL Likelihood = -0.22 Transmembrane 77 - 93 ( 76 - 93)

----- Final Results -----

5 bacterial membrane --- Certainty=0.1574(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 >GP:AAD33517 GB:AF132127 glucose-6-phosphate isomerase  
[Streptococcus mutans]  
Identities = 369/449 (82%), Positives = 408/449 (90%)

Query: 1 MSHITFDYSKVLSEFAGQHEIDFLQGQVTEADKLLREGTGPSSDFLGWLDLPENYDKDEF 60  
M+HI FDYSKVL F HE+D++Q QVT AD+ LR+GTGPG++ GWL+LP+NYDK+EF  
15 Sbjct: 1 MTHIKFDYSKVLGKFLASHELDYIQMVTADEALRKGTGPGAEMTGWLNLPQNYDKEEF 60

Query: 61 ARILTAEEKIKADSEVLVVGIGGSYLGAKAAIDFLNHHFANLQTAKERKAPQILYAGNS 120  
ARI AAEKIK+DSEVLVVGIGGSYLG+AAIDFLN F NL+ +ERKAPQILYAGNS  
Sbjct: 61 ARIKKAEEKIKSDSEVLVVGIGGSYLGARAAIDFLNSSFVNLENKEERKAPQILYAGNS 120

20 Query: 121 ISSTYLADLVEYVQDKEFSVNVISKSGTTTEPAIAFRVFKELLVKKYQEEANKRIYATT 180  
ISS YLADLV+YV DK+FSVNVISKSGTTTEPAIAFRVFK+LLVKKYQEEAN+RIYATT  
Sbjct: 121 ISSNYLADLDYVADKDFSVNVISKSGTTTEPAIAFRVFKDLLVKKYQEEANQRIYATT 180

25 Query: 181 DKVKGAVKVEADANNWETFVVPDNGGRFSVLTAVGLLPAAASGADITALMEGANAARKD 240  
D+VKGAVKVEADAN WETFVVPD+VGGRF+VLTAVGLLPAAASGAD+ LM GA AAR+D  
Sbjct: 181 DRVKGAVKVEADANGWETFVVPDSVGGRTVLTAVGLLPAAASGADLDQLMAGAEARQD 240

30 Query: 241 LSSDKISENIAYQYAAVRNVLYRKGYITEILANYEPSLQYFGEWWKQLAGESEGKDQKGI 300  
SS ++SEN AYQYAA+RN+LYRKGY+TE+LANYEPSLQYF EWWKQLAGESEGKDQKGI  
Sbjct: 241 YSSAELSENEAYQYAAIRNLYRKGYVTEVLANYEPSLQYFSEWWKQLAGESEGKDQKGI 300

35 Query: 301 YPTSANFSTDLHSLGQFIQEGYRNLFTFVIRVDNPRKNVILPELAEDLDGLGYLQGDVD 360  
YPTSANFSTDLHSLGQFIQEG RNLFTFVIRV+ RKN+++PE AEDLDGL YLQGDVD  
Sbjct: 301 YPTSANFSTDLHSLGQFIQEGNRNLFETVIRVEKARKNVLPEAAEDLDGLAYLQGDVD 360

40 Query: 361 FVNKKATDGVLLAHTDGGVPMFVTLPAQDEFTLGYTIYFFELAIASGYMNAVNPFDQP 420  
FVNKKATDGVLLAHTDGGVPM F+T+P QDEFTLGY IYFFELAI +SGY+N VNPFDQP  
Sbjct: 361 FVNKKATDGVLLAHTDGGVPMFTLTIPEQDEFTLGYVIYFFELAIIGLSGYLNGVNPFDQP 420

Query: 421 GVEAYKRNMFALLGKPGFEALSaelNARL 449  
GVEAYK+NMFALLGKPGFE L AELNARL  
Sbjct: 421 GVEAYKRNMFALLGKPGFEELGAELNARL 449

45 The protein has homology with the following sequences in the databases:

>GP:CAB90755 GB:AJ400707 hypothetical protein [Streptococcus  
uberis]  
Identities = 58/91 (63%), Positives = 69/91 (75%)

50 Query: 6 KRYPTITIFLLGLTGLIFIAMQVVYGHLAGAQAIYQVGGMFGLLVKAMPDQLWRLVTPIF 65  
K P+T F L +T L+FI MQV YG A Q ++Q GGMFGL+VK+MP QLWRLVTPIF  
Sbjct: 5 KEKPVTFFFLSVTILLFIVMQVFYGSWAKSPQVVFQFGGMFGLVVKSMPSQLWRLVTPIF 64

Query: 66 IHIGFGHFFVNGLTLYFVGQIVEDLWGSRLF 96  
55 IHIG+ HF +N LTLYFVGQ+ E +WGSR F  
Sbjct: 65 IHIGWEHFLINSLTYFVGQLAESIWGSRRFF 95

An alignment of the GAS and GBS proteins is shown below:

Identities = 380/401 (94%), Positives = 392/401 (96%)

60 Query: 1 MDLPENYDKEEFSRIQKAAEEKIKSDSEVLVVGIGGSYLGAKAAIDFLNHHFANLQTAE 60  
+DLPENYDK+EF+RI AAEKIK+DSEVLVVGIGGSYLGAKAAIDFLN+HFANLQTA+E  
Sbjct: 49 LDLPENYDKDEFARILTAEEKIKADSEVLVVGIGGSYLGAKAAIDFLNHHFANLQTAE 108



-343-

Query: 61 RKAPQILYAGNSISSTYLADLVEYVQDKFESVNVISKSGTTTEPAIAFRVFKELLVKKYG 120  
 RKAPQILYAGNSISSTYLADLVEYVQDKFESVNVISKSGTTTEPAIAFRVFKELLVKKYG  
 Sbjct: 109 RKAPQILYAGNSISSTYLADLVEYVQDKFESVNVISKSGTTTEPAIAFRVFKELLVKKYG 168

Query: 121 QEEANKRIYATTDKVKGA VKVEADANNWETFVVPDNGGRFSLTAVGLLPAAAGADIT 180  
 QEEANKRIYATTDKVKGA VKVEADANNWETFVVPDNGGRFSLTAVGLLPAAAGADIT  
 Sbjct: 169 QEEANKRIYATTDKVKGA VKVEADANNWETFVVPDNGGRFSLTAVGLLPAAAGADIT 228

Query: 181 ALMEGANAARKDLSSDKISENIAYQYAAVRNVLYRKGYITEILANYEPSLQYFGGEWWKQL 240  
 ALMEGANAARKDLSSDKISENIAYQYAAVRNVLYRKGYITEILANYEPSLQYFGGEWWKQL  
 Sbjct: 229 ALMEGANAARKDLSSDKISENIAYQYAAVRNVLYRKGYITEILANYEPSLQYFGGEWWKQL 288

Query: 241 AGESEKGDQKGIYPTSANFSTDHLHSLGQFIQEGYRNLFFETV+RV+ PRKNV IPEL EDL 300  
 AGESEKGDQKGIYPTSANFSTDHLHSLGQFIQEGYRNLFFETV+RV+ PRKNV IPEL EDL  
 Sbjct: 289 AGESEKGDQKGIYPTSANFSTDHLHSLGQFIQEGYRNLFFETVIRVDNPRKNV IPELAEDL 348

Query: 301 DGLGYLQGGKDVDFVNKKATDGVLLAHTDGGVPMFVTLPTQDAYTLGYTIYFFELAI GLS 360  
 DGLGYLQGGKDVDFVNKKATDGVLLAHTDGGVPMFVTLPTQDAYTLGYTIYFFELAI +S  
 Sbjct: 349 DGLGYLQGGKDVDFVNKKATDGVLLAHTDGGVPMFVTLPTQDEFTLGYTIYFFELAI AVS 408

Query: 361 GYLNSVNPFDQPGVEAYKRNMFALLGKPGFEELS AELNARL 401  
 GY+N+VNPFDQPGVEAYKRNMFALLGKPGFE LS AELNARL  
 Sbjct: 409 GYMNAVNPFDQPGVEAYKRNMFALLGKPGFEALS AELNARL 449

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 273

A DNA sequence (GBSx0298) was identified in *S. agalactiae* <SEQ ID 873> which encodes the amino acid sequence <SEQ ID 874>. Analysis of this protein sequence reveals the following:

Possible site: 38  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -2.66 Transmembrane 654 - 670 ( 653 - 671)  
 INTEGRAL Likelihood = -1.65 Transmembrane 113 - 129 ( 113 - 129)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2062 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9463> which encodes amino acid sequence <SEQ ID 9464> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA81906 GB:U04863 alcohol dehydrogenase 2 [Entamoeba histolytica]  
 Identities = 536/864 (62%), Positives = 663/864 (76%), Gaps = 3/864 (0%)  
 Query: 20 ETTDVALAIDTLVQNGLKALDEMR--QLNQEQVDYIVAKASVAALDAHGELALHAVEETG 77  
 +T V I+ LV+ AL E + QE++DYIV KASVAALD H LA AVEETG  
 Sbjct: 5 QTMVDEHINQLVRKAQVALKEYLKPEYTQEKIDYIVKKASVAALDQHCALAAA AVEETG 64  
 Query: 78 RGVFEDKATKNLFACEHVVNMNRHTKTGVIEEDDVTGLTLIAEPVGVVCGITPTTNPTS 137  
 RG+FEDKATKN+FACEHV + MRH KTVG+I D + G+T IAEPVGVVCG+TP TNPTS  
 Sbjct: 65 RGIFEDKATKNIFACEHVTHEMRHAKTVGIINVDPLYGITEIAEPVGVVCGVTPVNTNPTS 124  
 Query: 138 TAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAAIAAGAPENCQWIEQPSIDAT 197  
 TAIFKSLIS+KTRNPI+F+FHPSA + S AA+IVRDAAIAAGAPENC+QWIE I+A+  
 Sbjct: 125 TAIFKSLISIKTRNPIVFSFHPSALKCSIMAAKIVRDAAIAAGAPENCQWIEFGGIEAS 184  
 Query: 198 NALMNHDGIATILATGGNAMVKAAYS CGKPALGVGAGNVPAYVEKSANIRQAAHDIVMSK 257  
 N LMNH G+ATILATGGNAMVKAAYS GKPALGVGAGNVP Y+EK+ NI+QAA+D+VMSK

-344-

Sbjct: 185 NKL MNHPGVATILATGGNAMVKAAYSSGK PALGVGAGNVPTIYEKTCNIKQAANDVVM SK 244

Query: 258 SFDNGMVCASEQAVIIDKEIYKEFVEEFKSYHTYFVNKKEKALLEEFCFGAKANSKNCAG 317  
SFDNGM+CASEQA IIDKEIY + VEE K+ YF+N++EKA LE+F FG A S +

5 Sbjct: 245 SFDNGMICASEQAIIIDKEIYDQVVEEMKTLGAYFINEEAKLEKFMFGVNAYSADVNN 304

Query: 318 AKLNPNIIVGKSAVWIAEQAGFTVPEGTNIIAAECTEVSEKEPLTREKLSPVIAVLKAE ST 377  
A+LNP G S W AEQ G VPE NI+ A C EV EPLTREKLSPV+A+LKAE+T

10 Sbjct: 305 ARLNPKCPGMS PQWFABEQVGIVKPEDCNII CAVCKEVGPNEPLTREKLSPVLA ILKAENT 364

Query: 378 EDGVEKARQMVEFNGLGHSAAIHTKADLAREFGTRIRAIRVIWNSPSTFGGIGDVYN AF 437  
+DG++KA MVEFNG GHSAAIH+ D + ++ ++A R++ N+PS+ GGIG +YN

Sbjct: 365 QDGIDKAEAMVEFN GRGHSAAIHSNDKAVVEKYALTMKACRILHNT PSSQGGIGSIYNI 424

15 Query: 438 LPSLTLCGSGYGRNSVGDNVSAINLLNIKKVGRRRNNMQWFKVPSKTYFERDSIQYLQKC 497  
PS TLGCGSYG NSV NV+ NLLNIK++ RRNN+QWF+VP K +FE SI+YL +

Sbjct: 425 WPSFTLTCGSGYGGNSVSANVTYHNNLLNIKLADRRNNLQWFRVPPKIFFEPHSIRYLAEL 484

20 Query: 498 RDVERVMIVTDHAMVELGSLDRIIEQLDLRRNKVVYQIFAEVEPD PDITVMKGTDL MRT 557  
+++ ++ IV+D M +LG++DR+++ L R N+V +IF +VEPDP I TV KG +M T

Sbjct: 485 KELS KIFIVSDRM MYKLGYVDRVMDVLKRRSNEVEIEIFIDVEPDPSIQTVQKGLAVMNT 544

Query: 558 FKPDTIIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKFP ELGKKTKFVA 617  
F PD IIA+GGGS MDAAK+MWL YE PE DF + QKF+D+RKRAFKFP +GKK + +

25 Sbjct: 545 FGPDNIIAIGGSSAMDAAKIMWLLYEHPEADFFAMKQKFIDLRKRAFKFP TMGKKARLIC 604

Query: 618 IPTTSGTGSEVTPFAVISDKANNRKYPIADYSLTPTVAIVDPALVMTVPGFIAADTGMDV 677  
IPTTSGTGSEVTPFAVISD +KYP+ADYSLTP+VAIVDP M++P ADTG+DV

30 Sbjct: 605 IPTTSGTGSEVTPFAVISDHETGKKYPLADYSLTPSVAIVDPMFTMSLPKRAIADTGLDV 664

Query: 678 LTHATEAYVSQMANDYTDGLALQAIKIVFDYLSVKDADFEAREKMHNA STMAGMAFAN 737  
L HATEAYVS MAN+YTDGLA +A+K+VF+ L +S + D EAREKMHNA+T+AGMAFA+

Sbjct: 665 LVHATEAYVSVMANEYTDGLAREAVKLVFENLLKSY-NGDLEAREKMHNAATIAGMAFAS 723

35 Query: 738 AFLGISHMAHKIGA QFHTVHGRTNAILLPYVIRYNGTRPAKTATWPKYNYRDEKYQD 797  
AFLG+ HSM AHK+GA FH HGR A+LLP+VIRYNG +P K A WPKYN+Y+AD++Y +

Sbjct: 724 AFLGMDHSM AHKVGAA FHLP HGRCAVALLPHVIRYNGQKPRKLAMWPKYNFYKADQRYME 783

40 Query: 798 IAKLLGLPAATPEEAVESYAKAVYDLGTRLGIKMNF RDQ GIDEKEWKEKSRELAFLAYED 857  
+A+++GL TP E VE++AKA +L F+ IDE W K E+A LA+ED

Sbjct: 784 LAQMVG LKCNTPAEGVEAF AKACEELMKATETITGFKKANIDEAAWMSKVP EMALLAFED 843

Query: 858 QCSPANPRLPMVDHMQEIIEDAYY 881  
QCSPANPR+PMV M++I++ AYY

45 Sbjct: 844 QCSPANPRVPMVKDMEKILKAAYY 867

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 875> which encodes the amino acid sequence <SEQ ID 876>. Analysis of this protein sequence reveals the following:

Possible site: 55

50 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -3.66	Transmembrane	643 - 659 ( 642 - 660)
INTEGRAL	Likelihood = -1.81	Transmembrane	102 - 118 ( 102 - 118)

55 ----- Final Results -----

bacterial membrane	---	Certainty=0.2466(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

60 The protein has homology with the following sequences in the databases:

>GP:AAA81906 GB:U04863 alcohol dehydrogenase 2 [Entamoeba histolytica]

Identities = 535/870 (61%), Positives = 669/870 (76%), Gaps = 3/870 (0%)

65 Query: 6 NTVETTSVSVTIDALVQKGLAAL EEMRKLD--QE QVDYIVAKASVAALDAHGELAKHAYE 63

-345-

+T +T +V I+ LV+K AL+E K + QE++DYIV KASVAALD H LA A E  
 Sbjct: 2 STQQTMTVDEHINQLVRKAQVALKEYLKPEYTQEKIDYIVKKASVAALDQHCALAAA AVE 61

5 Query: 64 ETGRGVFEDKATKHLFACEHVNNMRHQKTVGII EDDVTGLTLIAEPVGVICGITPTTN 123  
 ETGRG+FEDKATK++FACEHV + MRH KTVGII D + G+T IAEPVGV+CG+TP TN  
 Sbjct: 62 ETGRGIFEDKATKNIFACEHVTHEMRHAKTVGIINVDPLYGITEIAEPVGVVCGVTPVTN 121

10 Query: 124 PTSTAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAIAAGAPENCVQWVETPSL 183  
 PTSTAIFKSLIS+KTRNPI+F+FHPSA + S AA+IVRDAIAAGAPENC+QW+E +  
 Sbjct: 122 PTSTAIFKSLISIKTRNPVFSFHPALKCSIMAAKIVRDAIAAGAPENCIQWIEFGGI 181

15 Query: 184 EATNALMNHGDIATILATGGNAMVKAAYS CGKPALGVGAGNVPAYVEKSANIRQAHDIV 243  
 EA+N LMNH G+ATILATGGNAMVKAAYS GKPALGVGAGNVP Y+EK+ NI+QAA+D+V  
 Sbjct: 182 EASNKLMNHG VATILATGGNAMVKAAYSSGKPALGVGAGNVPYIEKTCNIKQAANDVV 241

20 Query: 244 MSKSFNDGMV CASEQAVIIDKEIYDDFVAEFKSYHTYFVNKKEKALLEEFCGAKANSKN 303  
 MSKSFNDGM+CASEQA IIDKEIYD V E K+ YF+N++EKA LE+F FG A S +  
 Sbjct: 242 MSKSFNDGMICASEQA A IIDKEIYDQVVEEMKTLGAYFINEEEKAKLEKFMFGVNAYSAD 301

25 Query: 304 CAGAKLNPNI V GK PATWIAEQAGFTVPEGTNILAAECKEVSENEPLTREKLSPVIAVLKS 363  
 A+LNP G W AEQ G VPE NI+ A CKEV NEPLTREKLSPV+A+LK+  
 Sbjct: 302 VNNARLNP KCPGMSQWFAEQVG I KVPEDCNIICAVCKE VGPNEPLTREKLSPVLA ILKA 361

30 Query: 364 ESREDGVEKARQMVEFNGLGHSAAIHTADAE LAKEFGTRIRAIRVIWNSPSTFGGIGDVY 423  
 E+ +DG++KA MVEFNG GHSAAIH+ D + +++ ++A R++ N+PS+ GGIG +Y  
 Sbjct: 362 ENTQDGDIDKAEAMVEFN GRGHSAAIHSNDKAVVEKYALTMACRILHNT PSSQGGIGSIY 421

35 Query: 424 NAFLPSLTLCGSGYGRNAVGDNVSAINLLNIKKVGRRRNNMQWFKVPSKTYFERDSIQYL 483  
 N PS TLGCGSYG N+V NV+ NLLNIK++ RRNN+QWF+VP K +FE SI+YL  
 Sbjct: 422 NYIWPSTL GCGSYGGSV SANVTYHNNLLNIKRLADRRNNLQWFRVPPKIFFEPHSIRYL 481

40 Query: 484 QKCRDVERVMIVTDHAMVELGFLDRIIEQLDLRRNKVVYQIFAEVDPDPDITVMKGTEL 543  
 + +++ ++ IV+D M +LG++DR+++ L R N+V +IF +VEPDP I TV KG +  
 Sbjct: 482 AELKELSKIFIVSDRMMYKLG YVDRVMDVLKRRSNEVEIEIFIDVEPDPSIQTVQKGLAV 541

45 Query: 544 MRTFKPDTIIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFM DIRKRAF KFPPELGKKTK 603  
 M TF PD IIA+GGGS MDAAK+MWL YE PE DF + QKF+D+RKRAF KFP +GKK +  
 Sbjct: 542 MNTFGPDNIIAIGGGSAMDAAKIMWLLYEHPEADFFAMKQKFIDLRKRAF KFPPTMGKKAR 601

50 Query: 604 FVAIPTTSGTGSEVTPFAVISDKANNRKYPIADYSLTPTVAIVDPALVLTVP GFIAADTG 663  
 + IPTTSGTGSEVTPFAVISD +KYP+ADYSLTP+VAIVDP +++P ADTG  
 Sbjct: 602 LICIPTTSGTGSEVTPFAVISDHETGKKYPLADYSLTPSVAIVDPMTMSLPKRAIADTG 661

55 Query: 664 MDVLTHATEAYVSQMANDFTDGLALQAIKIVFDNLEKSVKTADFEAREKMHNASTMAGMA 723  
 +DVL HATEAYVS MAN++TDGLA +A+K+VF+NL KS D EAREKMHNA+T+AGMA  
 Sbjct: 662 LDVLVHATEAYVSVMANEYTDGLAREAVKLVFENLLKSY-NGDLEAREKMHNAATIAGMA 720

60 Query: 724 FANAFLGISHSMAHKIGAQFHTVHGRNTAILLPYVIRYNGTRPAKTATWPKYNYYRADEK 783  
 FA+AFLG+ HMAHK+GA FH HGR A+LLP+VIRYNG +P K A WPKYN+Y+AD++  
 Sbjct: 721 FASAFLGMDHSM AHKVGAA FHLP HGRCAVALLPHVIRYNGQKPRKLAMWPKYNFYKADQR 780

65 Query: 784 YQDI AKLLGLPASTPPEAVESYAKAVYDLGCRVGIQMNFKAQGIDENEWKEHSRELAYLA 843  
 Y ++A+++GL +TP E VE++AKA +L FK IDE W E+A LA  
 Sbjct: 781 YMELAQMVGLKCNTPAEGVEAFKACEELMKATETTTGFKKANIDEAAWMSKVP EMALLA 840

70 Query: 844 YEDQCSPANPR LPMVDHMQEIIEDAYGYA 873  
 +EDQCSPANPR+PMV M++I++ AYY A  
 Sbjct: 841 FEDQCSPANPRVPMVKDMEKILKAAYYPIA 870

60 An alignment of the GAS and GBS proteins is shown below:

Identities = 827/880 (93%), Positives = 852/880 (95%)

65 Query: 12 MTEKTKAVETTDVALAIDTLVQNGLKALDEMRLNQEQVDYIVAKASVAALDAHGELALH 71  
 MTE VETT V++ ID LVQ GL AL+EMR+L+QEQVDYIVAKASVAALDAHGELA H  
 Sbjct: 1 MTEGHNTVETTSVSTIDALVQKGLAAL EEMRKLDQEQVDYIVAKASVAALDAHGELAKH 60

Query: 72 AVEETGRGVFEDKATKNLFACEHVNNMRHTKTGVIEEDDVTGLTLIAEPVGVVCGITP 131

-346-

A EETGRGVFEDKATK+LFACEHVNNMRH KTVG+IEEDDVTGLTLIAEPVGV+CGITP  
 Sbjct: 61 AYEETGRGVFEDKATKHLFACEHVNNMRHQKTVGIIIEEDDVTGLTLIAEPVGVICGITP 120

5 Query: 132 TTNPTSTAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAIAAGAPENCVQWIEQ 191  
 TTNPTSTAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAIAAGAPENCVQW+E  
 Sbjct: 121 TTNPTSTAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAIAAGAPENCVQWVET 180

10 Query: 192 PSIDATNALMNHGDIATILATGGNAMVKAAYSCGKPALGVGAGNVPAYVEKSANIRQAAH 251  
 PS++ATNALMNHGDIATILATGGNAMVKAAYSCGKPALGVGAGNVPAYVEKSANIRQAAH  
 Sbjct: 181 PSLEATNALMNHGDIATILATGGNAMVKAAYSCGKPALGVGAGNVPAYVEKSANIRQAAH 240

15 Query: 252 DIVMSKSFNGMVCASEQAVIIDKEIYKEFVEEFKSYHTYFVNKKEKALLEEFCFGAKAN 311  
 DIVMSKSFNGMVCASEQAVIIDKEIY +FV EFKSYHTYFVNKKEKALLEEFCFGAKAN  
 Sbjct: 241 DIVMSKSFNGMVCASEQAVIIDKEIYDDFVAEFKSYHTYFVNKKEKALLEEFCFGAKAN 300

Query: 312 SKNCAGAKLNPINIVGKSAVWIAEQAGFTVPEGTNILAAECTEVSEKEPLTREKLSPVIAV 371  
 SKNCAGAKLNPINIVGK A WIAEQAGFTVPEGTNILAAEC EVSE EPLTREKLSPVIAV  
 Sbjct: 301 SKNCAGAKLNPINIVGKPATWIAEQAGFTVPEGTNILAAECKEVSENEPLTREKLSPVIAV 360

20 Query: 372 LKAESTEDGVEKARQMVEFNGLGHSAAIHTKDADLAREFGTRIRAIRVIWNSPSTFGGIG 431  
 LK+ES EDGVEKARQMVEFNGLGHSAAIHT DA+LA+EFGTRIRAIRVIWNSPSTFGGIG  
 Sbjct: 361 LKSESREDGVEKARQMVEFNGLGHSAAIHTADAELAKEFGTRIRAIRVIWNSPSTFGGIG 420

25 Query: 432 DVYNAFLPSLTGCGSYGRNSVGDNVSAINLLNIKKVGRRRNNMQWFKVPSKTYFERDSI 491  
 DVYNAFLPSLTGCGSYGRN+VGDNVSAINLLNIKKVGRRRNNMQWFKVPSKTYFERDSI  
 Sbjct: 421 DVYNAFLPSLTGCGSYGRNAVGDNVSAINLLNIKKVGRRRNNMQWFKVPSKTYFERDSI 480

30 Query: 492 QYLQKCRDVERVMIVTDHAMVELGFLDRIIEQLDLRRNKVVYQIFAEVEPDPTITVMKG 551  
 QYLQKCRDVERVMIVTDHAMVELGFLDRIIEQLDLRRNKVVYQIFAEVEPDPTITVMKG  
 Sbjct: 481 QYLQKCRDVERVMIVTDHAMVELGFLDRIIEQLDLRRNKVVYQIFAEVEPDPTITVMKG 540

35 Query: 552 TDLMRTFKPDITIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKFPELGK 611  
 T+LMRTFKPDITIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKFPELGK  
 Sbjct: 541 TELMRTFKPDITIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKFPELGK 600

Query: 612 KTKFVAIPTTSGTGSEVTPFAVISDKANNRKYPIADYSLTPTVAIVDPALVMTVPGFIAA 671  
 KTKFVAIPTTSGTGSEVTPFAVISDKANNRKYPIADYSLTPTVAIVDPALV+TVPGFIAA  
 Sbjct: 601 KTKFVAIPTTSGTGSEVTPFAVISDKANNRKYPIADYSLTPTVAIVDPALVLTVPGFIAA 660

40 Query: 672 DTGMDVLTHATEAYVSQMANDYTDGLALQAIKIVFDYLSVSKDADFEAREKMHNASTMA 731  
 DTGMDVLTHATEAYVSQMAND+TDGLALQAIKIVFD LE+SVK ADFEAREKMHNASTMA  
 Sbjct: 661 DTGMDVLTHATEAYVSQMANDFTDGLALQAIKIVFDNLEKSVKTADFEAREKMHNASTMA 720

45 Query: 732 GMAFANAFLGISHSMAHKIGAQFHTVHGRTNAILLPYVIRYNGTRPAKTATWPKYNYRA 791  
 GMAFANAFLGISHSMAHKIGAQFHTVHGRTNAILLPYVIRYNGTRPAKTATWPKYNYRA  
 Sbjct: 721 GMAFANAFLGISHSMAHKIGAQFHTVHGRTNAILLPYVIRYNGTRPAKTATWPKYNYRA 780

50 Query: 792 DEKYQDIAKLLGLPAATPEEAVESYAKAVYDLGTRLGIKMNFQDQIDEKEWKEKSRELA 851  
 DEKYQDIAKLLGLPA+TPEEAVESYAKAVYDLG R+GI+MNF+ QGIDE EWKE SRELA  
 Sbjct: 781 DEKYQDIAKLLGLPASTPEEAVESYAKAVYDLGCRVGIQMNFKAQGIDENEWKEKSRELA 840

Query: 852 FLAYEDQCSPANRPLPMVDHMQEIIEDAYYGYEERPGRRK 891  
 +LAYEDQCSPANRPLPMVDHMQEIIEDAYYGY ERPGRRK  
 Sbjct: 841 YLAYEDQCSPANRPLPMVDHMQEIIEDAYYGYAERPGRRK 880

A related GBS gene <SEQ ID 8533> and protein <SEQ ID 8534> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10  
 McG: Discrim Score: -4.68  
 GvH: Signal Score (-7.5): -2.48  
 Possible site: 21  
 >>> Seems to have no N-terminal signal sequence  
 ALOM program count: 1 value: -2.66 threshold: 0.0  
 INTEGRAL Likelihood = -2.66 Transmembrane 100 - 116 ( 99 - 117)  
 PERIPHERAL Likelihood = 3.61 173  
 modified ALOM score: 1.03

\*\*\* Reasoning Step: 3

----- Final Results -----

5           bacterial membrane --- Certainty=0.2062(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

SEQ ID 8534 (GBS432) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell  
 10 extract is shown in Figure 173 (lane 5; MW 66kDa). It was also expressed in *E.coli* as a His-fusion product.  
 SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 7; MW 41kDa).

GBS432-GST was purified as shown in Figure 223, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 vaccines or diagnostics.

### 15   **Example 274**

A DNA sequence (GBSx0299) was identified in *S.agalactiae* <SEQ ID 877> which encodes the amino acid  
 sequence <SEQ ID 878>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have no N-terminal signal sequence

20

----- Final Results -----

          bacterial cytoplasm --- Certainty=0.3444(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25

The protein has no significant homology with any sequences in the GENPEPT database, but there is  
 homology to SEQ ID 880.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for  
 vaccines or diagnostics.

### 30   **Example 275**

A DNA sequence (GBSx0300) was identified in *S.agalactiae* <SEQ ID 881> which encodes the amino acid  
 sequence <SEQ ID 882>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have a cleavable N-term signal seq.

35

INTEGRAL	Likelihood = -8.39	Transmembrane	74 - 90 ( 69 - 94)
INTEGRAL	Likelihood = -5.31	Transmembrane	168 - 184 ( 163 - 186)
INTEGRAL	Likelihood = -4.83	Transmembrane	34 - 50 ( 29 - 52)
INTEGRAL	Likelihood = -0.75	Transmembrane	202 - 218 ( 202 - 219)

40

----- Final Results -----

          bacterial membrane --- Certainty=0.4354(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45   The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA17305 GB:AL021926 hypothetical protein Rv0111 [Mycobacterium  
 tuberculosis]

Identities = 70/218 (32%), Positives = 104/218 (47%), Gaps = 12/218 (5%)

50

Query: 9   VRITGLLLVLLYHFFKNSFPGGFVGVDIFFTFSGFLITALLIDEFSKTKKIDFVVSFCRRR 68  
           +R   + LVL H           GGF+GVD FF SGFLIT+LL+DE +T +ID F RR

-348-

Sbjct: 39 LRAIAVALVLASHGGIPGMGGGFIGVDAFFVLSGFLITSLLLDELGRTGRIDLSGFWIRR 98

Query: 69 FYRIFPPLVLMVLVTIPFVFLVKSDFRASIGSQIMTALGFTSNFYEILTGGNYESQFI-P 127  
R+ P LVLMLV L + S + A +T+N+ + +Y+Q P

5 Sbjct: 99 ARRLPALVLMVLTVSAARALFPDQALTGLRSDAIAAFLWTANWRFVAQNTDYFTQGAPP 158

Query: 128 HLFVHTWSLSIEVHFYVLWGL---TVWLLSKRSKDQKQLRGTLFLISMGI FGVSFLTMF 183  
HTWSL +E +YV+W L LL+ R++ ++ R T+ + F ++ L

10 Sbjct: 159 SPLQHTWSLGVVEEQYYVWPLLLIGATLLLAARAR-RRRRATVGGVRFAAFLIASLGTM 217

Query: 184 VRAFFVDNFST-----IYFSTLSHIFPFFLGAMVATI 215  
A F++ IYF T + +G+ A +

Sbjct: 218 ASATAAVAFSTAATRDRIYFGTDTRAQALLIGSAAAAL 255

- 15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 879> which encodes the amino acid sequence <SEQ ID 880>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have an uncleavable N-term signal seq

20	INTEGRAL	Likelihood = -10.83	Transmembrane	325 - 341 ( 313 - 346)
	INTEGRAL	Likelihood = -9.29	Transmembrane	237 - 253 ( 234 - 258)
	INTEGRAL	Likelihood = -7.91	Transmembrane	166 - 182 ( 162 - 188)
	INTEGRAL	Likelihood = -6.10	Transmembrane	72 - 88 ( 68 - 92)
	INTEGRAL	Likelihood = -4.09	Transmembrane	264 - 280 ( 260 - 281)
	INTEGRAL	Likelihood = -2.87	Transmembrane	371 - 387 ( 370 - 390)
25	INTEGRAL	Likelihood = -2.66	Transmembrane	34 - 50 ( 32 - 50)
	INTEGRAL	Likelihood = -1.91	Transmembrane	3 - 19 ( 3 - 19)
	INTEGRAL	Likelihood = -0.85	Transmembrane	136 - 152 ( 136 - 154)

----- Final Results -----

30 bacterial membrane --- Certainty=0.5331(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

35 Identities = 167/226 (73%), Positives = 195/226 (85%)

Query: 1 MRIKWFSLVRLTGLLLVLLYHFFKNSFPGGFVGVDIFFTFSGFLITALLIDEFSKTKKID 60  
MRIKWFS VR+TGLLLVLLYHFFKN FPGGF+GVDIFFTFSG+LITALLIDE++K + ID

40 Sbjct: 1 MRIKWFSFVRVTGLLLVLLYHFFKNVFPGGFIGVDIFFTFSGYLITALLIDEYTKKESID 60

Query: 61 FVSFCRRRFYRIFPPLVLMVLVTIPFVFLVKSDFRASIGSQIMTALGFTSNFYEILTGGN 120  
+ F +RRFYRI PPLVLM+L+TIPF FL+K DF A+IGSQI LGFT+N YEILTG +

Sbjct: 61 IIGFLKRRFYRIVPPLVLMILLTIPFTFLIKKDFIANIGSQITAVLGFTTNIYEILTGSS 120

45 Query: 121 YESQFIPHLFVHTWSLSIEVHFYVLWGLTVWLLSKRSKDQKQLRGTLFLISMGI FGVSFL 180  
YESQFIPHLFVHTWSL+IEVHFY+ WG+ VWLL++R + QKQLRG LFLIS+GIF +SFL

Sbjct: 121 YESQFIPHLFVHTWSLAIEVHFYLFWGVFVWLLARRKETQKQLRGLLFLISLGIFAISFL 180

50 Query: 181 TMFVRAFFVDNFSTIYFSTLSHIFPFFLGAMVATISGIREITGRFK 226  
+MF+R+F NFS IYFS+LSH FPFLLGAM ATI+GI E T RF+

Sbjct: 181 SMFIRSFMTSNFSLIYFSSLSHSFPFFLGAMFATITGINETTFRFQ 226

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 55 Example 276

A DNA sequence (GBSx0302) was identified in *S.agalactiae* <SEQ ID 883> which encodes the amino acid sequence <SEQ ID 884>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database:

!GB:AE004818 hypothetical protein [Pseudomonas aerug...  
!GB:AE004818 hypothetical protein [Pseudomonas aerug...  
!GB:AE004818 hypothetical protein [Pseudomonas aerug...  
!GB:AE004818 hypothetical protein [Pseudomonas aerug...  
!GB:AE004818 hypothetical protein [Pseudomonas aerug...

10

>GP:AAG07403 GB:AE004818 hypothetical protein [Pseudomonas aeruginosa]  
Identities = 33/80 (41%), Positives = 50/80 (62%)

15

Query: 45 KYVGSIVNHHMTGKGKLT YENG DYYK GDFVNGVFEGKGT FVSVHGW SYTGDFKKGQPDGQ 104  
+Y G +V+ + G+G+L Y+NG +Y G F +G+ G GT+ G Y+G F G DGQ  
Sbjct: 39 RYRGELVDGRLEGQGR LDYDNGAWYAGRFEHGLLHGHGTWQ GADGSRYSGGFAAGLFDGQ 98

20

Query: 105 GRLNAK NKKVYKGT FKGQGIY 124  
GRL + VY+G F+QG++  
Sbjct: 99 GRLAMADGSVYQGGFRQGLF 118  
Identities = 31/91 (34%), Positives = 46/91 (50%), Gaps = 2/91 (2%)

25

Query: 34 QGVFSYDGGKIKYVGSIVNHHMTGKGKLT YENG DYYK GDFVNGVFEGKGT FVSVHGW SYT 93  
QG YD G Y G + + G G +G Y G F G+F+G+G G Y  
Sbjct: 52 QGR LDYDNGAW-YAGRFEHGLLHGHGTWQ GADGSRYSGGFAAGLFDGQGR LAMADGSVYQ 110

30

Query: 94 GDFKKGQPDGQGR LNAK NKKVYKGT FKGQGIY 124  
G F++G DG+G L + + Y+G F++G+Y  
Sbjct: 111 GGFQGLFDGEGSLEQQGTR-YRGGFRKGLY 140  
Identities = 31/91 (34%), Positives = 42/91 (46%), Gaps = 1/91 (1%)

35

Query: 32 SSQGVFSYDGGKIKYVGSIVNHHMTGKGKLT YENG DYYK GDFVNGVFEGKGT FVSVHGW 91  
S QG G +Y GS + G+G + G+ Y G F +G GKG + G  
Sbjct: 141 SGQGTLDGSDGS-RYQGSFRQGRLEGE GSFSDSQGNQYAGTFRDQLNGKGRWSGPDGDR 199

40

Query: 92 YTGDFKKGQPDGQGR LNAK NKKVYKGT FKGQ 122  
Y G FK Q GQGR + + V+ G F +G  
Sbjct: 200 YVGQFKDNQFHGQGRYESASGDVWIGRFSEG 230  
Identities = 31/91 (34%), Positives = 45/91 (49%), Gaps = 4/91 (4%)

45

Query: 34 QGVFSYDGGK----IKYVGSIVNHHMTGKGKLT YENG DYYK GDFVNGVFEGKGT FVSVHG 89  
QG+F +G +Y G +G+G L +G Y+G F G EG+G+F G  
Sbjct: 115 QGLFDGEGSLEQQGTRYRGGRKGLYSQGTL DGS DGSRYQGSFRQGRLEGE GSFSDSQG 174

50

Query: 90 WSYTGDFKKGQPDGQGR LNAK NKKVYKGT FKG 120  
Y G F+ GQ +G+GR + + Y G FK  
Sbjct: 175 NQYAGTFRDQLNGKGRWSGPDGDRYVGQFK 205  
Identities = 28/87 (32%), Positives = 45/87 (51%), Gaps = 1/87 (1%)

55

Query: 34 QGVFSYDGGKIKYVGSIVNHHMTGKGKLT YENG DYYK GDFVNGVFEGKGT FVSVHGW SYT 93  
+G FS G +Y G+ + + GKG+ + +GD Y G F + F G+G + S G +  
Sbjct: 166 EGSFSDSQGN-QYAGTFRDQLNGKGRWSGPDGDRYVGQFKDNQFHGQGRYESASGDVWI 224

60

Query: 94 GDFKKGQPDGQGR LNAK NKKVYKGT FKG 120  
G F +G +G G L + Y+G F+  
Sbjct: 225 GRFSEGALNGPGELLGADGSR YRGGFQ 251  
Identities = 28/89 (31%), Positives = 43/89 (47%), Gaps = 2/89 (2%)

65

Query: 34 QGVFSYDGGKIKYVGSIVNHHMTGKGKLT YENG DYYK GDFVNGVFEGKGT FVSVHGW SYT 93  
QG + G + Y G +G+G L + G Y+G F G++ G+GT G Y  
Sbjct: 98 QGR LAMADGSV-YQGGFRQGLFDGEGSLE-QQGTRYRGGRKGLYSQGTL DGS DGSRYQ 155

Query: 94 GDFKKGQPDGQGR LNAK NKKVYKGT FKGQ 122  
G F++G+ +G+G + Y GTF+ G  
Sbjct: 156 GSFRQGRLEGE GSFSDSQGNQYAGTFRDG 184

-350-

Identities = 25/80 (31%), Positives = 37/80 (46%)

Query: 45 KYVGSIVNHHMTGKGLTYENGDIYKGD FVNGVFEGKGT FVSVHGW SYTGDFKKGQPDGQ 104  
 +YVG ++ G+G+ +GD + G F G G G + G Y G F+ + GQ  
 Sbjct: 199 RYVGQFKDNQFHGQGRYESASGDVWIGRFSEGLNPGELLGADGSR YRGGFQFWRFHGQ 258

Query: 105 GRLNAKNNKKVYKGT FKQGIY 124  
 G L + Y+G F G Y  
 Sbjct: 259 GLLEQLDGTREYEGGFAAGAY 278

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 885> which encodes the amino acid sequence <SEQ ID 886>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -13.16 Transmembrane 20 - 36 ( 12 - 41)

----- Final Results -----

bacterial membrane --- Certainty=0.6265(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAA16606 GB:D90899 hypothetical protein [Synechocystis sp.]  
 Identities = 37/89 (41%), Positives = 49/89 (54%), Gaps = 6/89 (6%)

Query: 48 KGRMHYT-----GYVINHKMNGEGKLVYPNGDIYEGTFKDGLFEGKGTFTAKTGWLYNG 101  
 KG YT G V+ ++NG GK Y NGD YEGT K+G +G+G F G Y G  
 Sbjct: 141 KGTFIYTNGDRCSGT VVQGE LNSGKCEYNNNGDQYEGT LKNGQPDGEGIFRFAAGGEYEG 200

Query: 102 EFHKGQANGKGV LKAKNNKVYKGIFKQGI 130  
 EF G+ +G+G N ++G FKQG+  
 Sbjct: 201 EFQSGEFSGQGT RIFANGNR FQGQFKQGL 229

An alignment of the GAS and GBS proteins is shown below:

Identities = 68/126 (53%), Positives = 93/126 (72%)

Query: 1 MKNFKITRTHLEILSLIIIVFGLSVFTLTSSQGVFSYDGGKIKYVGSIVNHHMTGK GK 60  
 +K + ITR LEI+S+I+I+V +SVF++ S++ +YD G++ Y G ++NH M G+GK  
 Sbjct: 8 VKKWSITRAKLEIVSVIVILVCAISVFSVRISNKTSLTYDKGRMHYTGYVINHKMNGEGK 67

Query: 61 LTYENGDIYKGD FVNGVFEGKGT FVSVHGW SYTGDFKKGQPDGQRLNAKNNKVYKGT FK 120  
 L Y NGD Y+G F +G+FEGKGT F + GW Y G+F KGQ +G+G L AKN KVKYK FK  
 Sbjct: 68 LVYPNGDIYEGTFKDGLFEGKGTFTAKTGWLYNGEFHKGQANGKGV LKAKNNKVYKGIFK 127

Query: 121 QGIYQK 126  
 QGI+QK  
 Sbjct: 128 QGIFQK 133

SEQ ID 884 (GBS139) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 19 (lane 3; MW 13kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 22 (lane 2; MW 38.2kDa), in Figure 24 (lane 7; MW 38kDa) and in Figure 33 (lane 7; MW 38.2kDa).

The GBS139-GST fusion product was purified (Figure 200, lane 2) and used to immunise mice. The resulting antiserum was used for FACS (Figure 287), which confirmed that the protein is immunoaccessible on GBS bacteria.



-351-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 277**

A DNA sequence (GBSx0303) was identified in *S.agalactiae* <SEQ ID 887> which encodes the amino acid sequence <SEQ ID 888>. This protein is predicted to be holliday junction dna helicase ruvb (ruvB). Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.4386(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB75331 GB:Y15896 RuvB protein [Bacillus subtilis]

Identities = 196/322 (60%), Positives = 254/322 (78%)

20 Query: 3 RFLSDAMGDEELVERTLRPOYLREYIGQDKVKDQLKIFIEAAKLDES LDHVLVLF GPPG 62  
R + S+A E ++E++LRPQ L +YIGQ KVK+ L++FI+AAK+R E+LDHVLL+GPPG  
Sbjct: 4 RLVSSSEADNHESVIEQSLRPQNLAQYIGQHKVKENLRV FIDAAKMRQETLDHVLVLYGPPG 63

25 Query: 63 LGKTTMAFVIANELGVNLKQTS GPAIEKSGDLVAILNDLEPGDVL FIDEIHRMFMAVEEV 122  
LGKTT+A ++ANE+GV L+ TSGPAIE+ GDL AIL LEPGDVL FIDEIHR+ ++EEV  
Sbjct: 64 LGKTTLASIVANEMGVELRTTSGPAIERPGDLAAILTALEPGDVL FIDEIHR LRSIEEV 123

30 Query: 123 LYSAMEDFYIDIMIGAGETSRSVHLDLPFFTLIGATTAGMLSNPLRARFGITGHMEYYE 182  
LY AMEDF +DI+IG G ++RSV LDLPPFTL+GATTR G+L+ PLR RFG+ +EYY  
Sbjct: 124 LYPAMEDFCLDIVIGKGPSARSVRDLDPFFTLVGATTRVGLLTAPLRDRFGVMSRLEYTT 183

35 Query: 183 ENDLTEIIERTADIFEMKITYEASELARRSRGTPRIANRLLKRVRDYQAQIMGDLIDDN 242  
+ +L +I+ RTAD+FE++I +A E+ARRSRGTPR+ANRLL+RVRD+AQ++GD I ++  
Sbjct: 184 QEEELADIVTRTADVFEVEIDKPSALEIARRSRGTPRVANRLLRVRDFAQVLGDSRITED 243

40 Query: 243 ITDKALTMLDVDHEGLDYVDQKILRTMIEMYNGGPVGLGTLVSNIAEERDTVEDMYEPYL 302  
I+ AL L VD GLD++D K+L MIE +NGGPVGL T+S I EE T+ED+YEPYL  
Sbjct: 244 ISQNALERLQVDRGLDHDHKLMLGMIEKFNGGPVGLDTISATIGESHTIEDVYEPYL 303

45 Query: 303 IQKGFIMRTRTGRVATVKAYEH 324  
+Q GFI RT GR+ T Y H  
Sbjct: 304 LQIGFIQRTPRGRIVTPAVYHH 325

A related GBS nucleic acid sequence <SEQ ID 10943> which encodes amino acid sequence <SEQ ID 10944> was also identified.

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 889> which encodes the amino acid sequence <SEQ ID 890>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.0686(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 An alignment of the GAS and GBS proteins is shown below:

Identities = 282/327 (86%), Positives = 306/327 (93%)

-352-

Query: 1 MTRFLDSAMGDEELVERTLRPQYLREYIGQDKVKDQLKIFIEAAKLDESLDHVLLFGP 60  
 M R LD++ MG+EE +RTLRPQYL EYIGQDKVK+Q IFIEAAK RDESLDHVLLFGP  
 Sbjct: 25 MARILDNNVMGNEEFSDRTLRPQYLHEYIGQDKVKEQFAIFIEAAKRDESLDHVLLFGP 84

5 Query: 61 PGLGKTTMAFVIANELGVNLKQTS GPAIEKSGDLVAILNDLEPGDVL FIDEIHRMPMAVE 120  
 PGLGKTTMAFVIANELGVNLKQTS GPA+EK+GDLVAILN+LEPGD+LFIDEIHRMPM+VE  
 Sbjct: 85 PGLGKTTMAFVIANELGVNLKQTS GPAVEKAGDLVAILNELEPGDIL FIDEIHRMPMSVE 144

10 Query: 121 EVLYSAMEDFYIDIMIGAGETSRVHLDLPFFTLLIGATTAGMLSNPLRARFGITGHMEY 180  
 EVLYSAMEDFYIDIMIGAG+TSRS+HLDLPFFTLLIGATTAGMLSNPLRARFGITGHMEY  
 Sbjct: 145 EVLYSAMEDFYIDIMIGAGDTSRSIHLDLPPFFTLLIGATTAGMLSNPLRARFGITGHMEY 204

15 Query: 181 YEENDLTEIIERTADIFEMKITYEASELARRSRGTPRIANRLLKVRDYAQIMGDGLID 240  
 Y+E DLTEI+ERTA IFE+KI +EAA +LA RSRGTPRIANRLLKVRDYAQI+GDG+I  
 Sbjct: 205 YQEKDLTEIVERTATIFEIKIDHEAARKLACRSRGTPRIANRLLKVRDYAQIIGDGIIT 264

20 Query: 241 DNITDKALTMLDVDHEGLDYVDQKILRTMIEMYNGGPVGLGTL SVNIAEERDTVEDMYEP 300  
 ITD+ALTMLDVD EGLDY+DQKILRTMIEMY GGPVGLGTL SVNIAEER+TVE+MYEP  
 Sbjct: 265 AQITDRALTMLDVDREGLDYIDQKILRTMIEMYQGGPVGLGTL SVNIAEERNVTVEEMYEP 324

Query: 301 YLIQKGFIMRTRTGRVATVKAYEHLGY 327  
 YLIQKGF+MRTRTGRVAT KAY HLG Y  
 Sbjct: 325 YLIQKGFIMRTRTGRVATQKAYRHLGY 351

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 278

A DNA sequence (GBSx0304) was identified in *S. agalactiae* <SEQ ID 891> which encodes the amino acid sequence <SEQ ID 892>. Analysis of this protein sequence reveals the following:

Possible site: 43  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -2.87 Transmembrane 157 - 173 ( 157 - 174)  
 INTEGRAL Likelihood = -1.49 Transmembrane 205 - 221 ( 205 - 222)

----- Final Results -----  
 bacterial membrane --- Certainty=0.2147(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 893> which encodes the amino acid sequence <SEQ ID 894>. Analysis of this protein sequence reveals the following:

Possible site: 56  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3097(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 130/303 (42%), Positives = 202/303 (65%)

Query: 1 MLKHFGSKVRNLRVTRNITREDFCGDETELSVRQLARIESGQSIPNLTKAHYIAKQLNVK 60  
 ML+HFG KV+ LR+ + I+RED CGDE+ELSVRQLARIE GQSIP+L+K +IAK LNV  
 Sbjct: 1 MLEHFGGKVKVLRLEKRISREDLCGDESELSVRQLARIELGQSIPSLSKVIFIAKALNV 60

Query: 61 LDILTGGESLELPKRYKELKYILIRIPTYADAERLKLRECQFDHIFEFFYDNLPEDECLA 120  
 + LT G LELPKRYKELKYILIR PTY D +L++RE QFD IFE++YD LPE+E +

-353-

Sbjct: 61 VGYLTDGADLELPKRYKELKYLILRTPYMDGKLQVREEQFDEIFEDYDKLPPEEKII 120

Query: 121 IDSLQAKFEVYQTGDINFGVEVLCECFDKVKYKEKYTLNDLIIIDFLTCVAVSKFNNRA 180  
ID LQA + + + NFG+++L E F+++K K ++ NDLI+++L+L + + +

5 Sbjct: 121 IDCLQATLDTLLSENTNFGIDLLQEFYFNQIKTKVRFQRNDLILLELYLAYLDIEGMDGQY 180

Query: 181 FTKEVFQTICTKLISQNHKLTAEDLEFWFNHVLNLCVFGVGLCLNSEECLEMLEVSRQTMV 240  
K + ++ L Q + ++LF N +++ + L N + L + +E+S++ M

10 Sbjct: 181 SDKIFYDSLNDLSEQFEQFELDELFIVNKIIIDISSLSLKNRLDNLEKAIEMSQKIMA 240

Query: 241 STHDFHKMPLYFMYQWKYFITDNDIKSAENAYQQSIMFSKMIDDKHLIKKLELEWQEDI 300  
D+++MP+ + +WKYF+ DI AE ++ ++ +F++M D++L KL EW++D+

Sbjct: 241 KIQDWNRMPIKLIEWKYFLIKQDKIIEAQSFMAKCLFAQMTADQYLENKLIQEWKDV 300

15 Query: 301 TGH 303  
+  
Sbjct: 301 KSY 303

SEQ ID 892 (GBS319) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell  
20 extract is shown in Figure 40 (lane 4; MW 37kDa). It was also expressed in *E.coli* as a GST-fusion  
product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 7; MW 62kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
vaccines or diagnostics.

### Example 279

25 A DNA sequence (GBSx0305) was identified in *S.agalactiae* <SEQ ID 895> which encodes the amino acid  
sequence <SEQ ID 896>. This protein is predicted to be adenylosuccinate lyase (purB). Analysis of this  
protein sequence reveals the following:

Possible site: 35  
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.3358(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04344 GB:AP001509 adenylosuccinate lyase [Bacillus halodurans]  
Identities = 326/430 (75%), Positives = 366/430 (84%)

40 Query: 1 MIERYSRPEMAAIWTEENKYRAWLEVEILADEAWAELGEIPKEDVAKIREKADFDIDRIL 60  
MIERY+RPEM AIWTEEN+Y+AWLEVEI+A EAWAELGEIPKEDV KIRE A FD++RIL  
Sbjct: 1 MIERYTRPEMGAIWTEENRYQAWLEVEIVACEAWAELGEIPKEDVKKIREHASFDVERIL 60

45 Query: 61 EIEQDTRHDVVAFTRAVSETLGEERKWWHYGLTSTDVVDYAYGYLYKQANDIIRRDLENF 120  
EIEQ+TRHDVVAFTRAVSETLGEERKWWHYGLTSTDVVDY YL KQAN+II DL F  
Sbjct: 61 EIEQETRHDVVAFTRAVSETLGEERKWWHYGLTSTDVVDYALSYYLLKQANEIIEADLVRF 120

Query: 121 TNIVADKAKEHKFTIMGRTHGVHAEPTTFGLKLATWYSEMKNRNERFEHAAAGVEAGKI 180  
+I+ +KA EHK+T+MMGRTHGVHAEPTTFGLKLA WY EMKRN+ERF AA GV GK+

50 Sbjct: 121 LDILKEKALEHKYTVMMGRTHGVHAEPTTFGLKLALWYEEEMKRNLERFRLAAEGVRVVGKL 180

Query: 181 SGAVGNFANIPPFVEQYVCDKLGIRPQEISTQVLPRLDHAIFYFAVLASIATSIERMATEI 240  
SGAVG +ANI PFVEQYVC+KLG+ ISTQ L RD HAEY A LA IATSIE+ A EI

Sbjct: 181 SGAVGTYANIDPFVEQYVCEKGLERAPISTQTLQRDRHAEYMATLALIATSIEKFAVEI 240

55 Query: 241 RGLQKSEQREVEEFFAKGQKGSSAMPHKRNPIGSENMTGLARVIRGHMVTAYENVALWHE 300  
RGLQKSE REVEE+FAKGQKGSSAMPHKRNPIGSENMTG+ARV+RGHM+ AYENV LWHE  
Sbjct: 241 RGLQKSETREVEEYFAKGQKGSSAMPHKRNPIGSENMTGIARVVIRGHMLAAYENVPLWHE 300

-354-

Query: 301 RDISHSSAERIITPDTTILIDYMLNRFNIVKNLTVFPENMMRNMESTFGLIYSQVRMLK 360  
 RDISHSSAERII PD TI I+YMLNRFNIVKNLTVFPENM RNM T+GLIYSQRV+L  
 Sbjct: 301 RDISHSSAERIILPDATIAINYMLNRFNIVKNLTVFPENMKRNMTRTYGLIYSQRVLLS 360

5 Query: 361 LIEKGMTREEAYDLVQPKTAYSWDNQVDFKPLLEEDTKVTSCLTQEEIDELFNPIYYTKR 420  
 LI+KGM REEAYDLVQPK +W+ V F+ L+E++ ++TS L+ EEI+ F+ ++ K  
 Sbjct: 361 LIDKGMVREEAYDLVQPKAMEAWKGVQFRELVEQEERITSVLSPEEIEACFDYNHHLKH 420

10 Query: 421 VDDIFERLGL 430  
 VD IFERLGL  
 Sbjct: 421 VDTIFERLGL 430

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 897> which encodes the amino acid sequence <SEQ ID 898>. Analysis of this protein sequence reveals the following:

15 Possible site: 35  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.3358(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 422/430 (98%), Positives = 428/430 (99%)

25 Query: 1 MIERYSRPEMAAIWTEENKYRAWLEVEILADEAWAELGEIPKEDVAKIREKADFDIDRIL 60  
 M+ERYSRPEMAAIWTEENKY AWLEVEILADEAWAELGEIPKEDVAKIREKADFDIDRIL  
 Sbjct: 1 MLERYSRPEMAAIWTEENKYHAWLEVEILADEAWAELGEIPKEDVAKIREKADFDIDRIL 60

30 Query: 61 EIEQDTRHDVVAFTRAVSETLGEERKWHVHYGLTSTDVVDYAGYLYKQANDIIRRDLNLF 120  
 EIEQDTRHDVVAFTRAVSETLGEERKWHVHYGLTSTDVVDYAGYLYKQANDIIRRDLNLF  
 Sbjct: 61 EIEQDTRHDVVAFTRAVSETLGEERKWHVHYGLTSTDVVDYAGYLYKQANDIIRRDLNLF 120

35 Query: 121 TNIVADKAKEHKFTIMMGRTHGVHAEPTTFGLKLATWYSEMKNIERFEHAAAGVEAGKI 180  
 TNIVADKA+EHK TIMMGRTHGVHAEPTTFGLKLATWYSEMKNIERFEHAAAGVEAGKI  
 Sbjct: 121 TNIVADKAREHKMTIMMGRTHGVHAEPTTFGLKLATWYSEMKNIERFEHAAAGVEAGKI 180

40 Query: 181 SGAVGNFANIPPFVEQYVCDKLGIRPQEISTQVLPRDLHAEYFAVLASIATSIERMATEI 240  
 SGAVGNFANIPPFVE+YVCDKLGIRPQEISTQVLPRDLHAEYFAVLASIATSIERMATEI  
 Sbjct: 181 SGAVGNFANIPPFVEEYVCDKLGIRPQEISTQVLPRDLHAEYFAVLASIATSIERMATEI 240

45 Query: 241 RGLQKSEQREVEEFFAKGQKGSSAMPHKRNPIGSENMTGLARVIRGHMVTAYENVALWHE 300  
 RGLQKSEQREVEEFFAKGQKGSSAMPHKRNPIGSENMTGLARVIRGHMVTAYENV+LWHE  
 Sbjct: 241 RGLQKSEQREVEEFFAKGQKGSSAMPHKRNPIGSENMTGLARVIRGHMVTAYENVSLWHE 300

50 Query: 301 RDISHSSAERIITPDTTILIDYMLNRFNIVKNLTVFPENMMRNMESTFGLIYSQVRMLK 360  
 RDISHSSAERIITPDTTILIDYMLNRFNIVKNLTVFPENMMRNMESTFGLIYSQVRMLK  
 Sbjct: 301 RDISHSSAERIITPDTTILIDYMLNRFNIVKNLTVFPENMMRNMESTFGLIYSQVRMLK 360

55 Query: 361 LIEKGMTREEAYDLVQPKTAYSWDNQVDFKPLLEEDTKVTSCLTQEEIDELFNPIYYTKR 420  
 LIEKGMTREEAYDLVQPKTAYSWDNQVDFKPLLEEDTKVTSCLTQEEIDELFNPIYYTKR  
 Sbjct: 361 LIEKGMTREEAYDLVQPKTAYSWDNQVDFKPLLEEDTKVTSCLTQEEIDELFNPIYYTKR 420

Query: 421 VDDIFERLGL 430  
 VDDIF+RLG+  
 Sbjct: 421 VDDIFKRLGI 430

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-355-

**Example 280**

A DNA sequence (GBSx0306) was identified in *S.agalactiae* <SEQ ID 899> which encodes the amino acid sequence <SEQ ID 900>. Analysis of this protein sequence reveals the following:

Possible site: 45

5 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -16.24	Transmembrane	145 - 161 ( 119 - 167)
INTEGRAL	Likelihood = -9.98	Transmembrane	125 - 141 ( 119 - 144)
INTEGRAL	Likelihood = -9.29	Transmembrane	28 - 44 ( 23 - 51)
INTEGRAL	Likelihood = -7.01	Transmembrane	196 - 212 ( 193 - 220)
10 INTEGRAL	Likelihood = -6.21	Transmembrane	96 - 112 ( 88 - 116)
INTEGRAL	Likelihood = -5.79	Transmembrane	249 - 265 ( 246 - 266)
INTEGRAL	Likelihood = -2.87	Transmembrane	222 - 238 ( 222 - 238)
INTEGRAL	Likelihood = -2.28	Transmembrane	279 - 295 ( 278 - 295)

15 ----- Final Results -----

bacterial membrane	---	Certainty=0.7496(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

20 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB13498 GB:AB028634 RNA polymerase [Flammulina velutipes]

Identities = 83/336 (24%), Positives = 150/336 (43%), Gaps = 40/336 (11%)

25 Query: 152 ILLLI AFV SIGKNR-VYNFVQNLNYFEEVIWNYFEENPVKIKEKSLIIK-----FLLTIS 205  
IL L SI NR + ++ N ++ N+F+ + +K K L+I F++ +S  
Sbjct: 133 ILFLYLYISILINRFILKWL DNSGIYKININWFKNHMIKHINKMLVINIKFFNFIIKLS 192

30 Query: 206 FVFVIDFAMVRL-----LNFNIKFS TILACSAILLAWLYQN-----KSVTEPF 249  
+ +I +++ L +NF+I+ I I ++ S+ F  
Sbjct: 193 IITIIGISIME LFGIFGINFDIRIIINYLKTINSKG IHLTIINMDQYSVLENSIHTIFY 252

35 Query: 250 LKKLVIYFIFFIATLIGNLKN-ELSILETPLL FISISFFTMDRIALSKEMRDLI--ISKS 306  
+ L+I+ IF L N+KN + +I +L+I IF I ++DL+ ++K  
Sbjct: 253 INLLIIFLIFISLILYRNVNIDTNIKRWIIYILIFLINIIFIFNHIYIKDLMDNLNKY 312

40 Query: 307 ILFYD HENIKPSILLSEIKEIKYLENVDIGE---LELV RQMVIRLRLELEEEFLILSDI 363  
IL Y D I S+ L ++K L+ ++I + V+ + I+ ++E L + I  
Sbjct: 313 ILDYMDLHIIVNSLFLFNKFDVK-LKRINIYKSYSTVTVKDLEIKSKIEERSNELDIKLI 371

45 Query: 364 YMKNG-YEKYIQFVQGNVYFINLE--LDKIPNYTNLKLILESIFD---HNNQKIFIPKL 416  
K G YE YI ++ N+ ++ E L P Y N +E + + + F+ K+  
Sbjct: 372 IAKYGSYENYINSIE-NINIVDEEFILKNYPYINDSKFIEFLMELEPLFRDHTEFVKKI 430

Query: 417 YEEYIYILISLGEVEKAREIL---KEVSDYLTEESL 449  
YE L + K+IL KE+ DY+ + +L  
Sbjct: 431 YENLNSTNEKLEFLLANKDILSENKEIFDYVLQLNL 466

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 281**

A DNA sequence (GBSx0308) was identified in *S.agalactiae* <SEQ ID 901> which encodes the amino acid sequence <SEQ ID 902>. Analysis of this protein sequence reveals the following:

Possible site: 37

55 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.3307(Affirmative)	< succ>
---------------------	-----	-------------------------------	---------

-356-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 282

10 A DNA sequence (GBSx0309) was identified in *S.agalactiae* <SEQ ID 903> which encodes the amino acid sequence <SEQ ID 904>. This protein is predicted to be purK (purK). Analysis of this protein sequence reveals the following:

Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.0334 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

20 A related GBS nucleic acid sequence <SEQ ID 9461> which encodes amino acid sequence <SEQ ID 9462> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

25 >GP:CAA04376 GB:AJ000883 purK [Lactococcus lactis]  
 Identities = 208/347 (59%), Positives = 258/347 (73%), Gaps = 3/347 (0%)  
 Query: 14 NSFKTIGIIGGGQLGQMMAIAAIYMGHKVITLDPASDCPASRVS-EVIVAPYDDVEALGT 72  
           N+ +TIGIIGGGQLGQMMAIAA YMGHKVITLDP +C A++VS E+IVAPYDDVE L  
 Sbjct: 4 NTKQTIGIIGGGQLGQMMAIAAQYMGHKVITLDPNPNCSAAKVSDELIVAPYDDVENLLR 63  
 30 Query: 73 LAARCDVLTIEFENVADGLDAVVSAGQLPQGTDLRLISQNRIFEKDFLANKAGVTVAPY 132  
           LA CDV+TYEFENV A L + ++PQG LL I+QNR FEK+FL N+A V VAP+  
 Sbjct: 64 LAYACDVITYEFENVSAKALHEIEGCVRIQQGIRLLEITQNRFFEKEFLTNEAKVNVAPW 123  
 35 Query: 133 KVTSSLDLEGLDLTKTYVLKTATGGYDGHGQKVIKSAEDLPEAQQLANSAQCVEEFVN 192  
           ++V S+ L +T+ VLKT TGGYDGHGQ V+ + E L A+ L ++CVLE+F++  
 Sbjct: 124 QLVDSAEKLPEP+VTRKQVLKTTTGGYDGHGQVVLNTDEKLSAAKSLTELSECVLEDFIS 182  
 40 Query: 193 FDLEISVIVSGNGQDVTVPVQENIHRNLSKTIVPARISDQLADKAKEMAVQIAKKLQ 252  
           F+ EISVI+SGNG + VFP+ EN HR NIL +TI PARIS ++ + A ++A IA+KL+  
 Sbjct: 183 FERISVLIISGNGHEYVVFPLAENEHRENILHQTISPARISEITENAYKIATSIAEKL 242  
 45 Query: 253 LSGTLCVEMFATAD-DIIVNEIAPRPHNSGHYSIEACDFSQFDTHILGVLGAPLPPIKLH 311  
           LSG LCVEMF TAD I VNE+APRPHNSGH++IEACDF+QFD HI G+LG LP KL  
 Sbjct: 243 LSGVLCVEMFLTADGQIYVNELAPRPHNSGHFTIEACDFNQFDLHIKGI LGEDLPEPKLL 302  
 Query: 312 APAVMFNVLGQHVQQAIDHVAQNPSAHLHMYGKLEAKHNRKMGHVTV 358  
           PA+M NVLGQHV+ ++ H H YGK +AKHNRKMGHVT+  
 Sbjct: 303 KPAIMLVNLGQHVEAVKKLNHEHADWHQHDYKADAKHNRKMGHVTI 349

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 905> which encodes the amino acid sequence <SEQ ID 906>. Analysis of this protein sequence reveals the following:

Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----

-357-

bacterial cytoplasm --- Certainty=0.0334(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

## 5 An alignment of the GAS and GBS proteins is shown below:

Identities = 344/369 (93%), Positives = 353/369 (95%)

Query: 1 MRNKEKSQRSQAMNSFKTIGIIGGGQLGQMMIAAIIYMGHKVITLDPASDCPASRVSEVI 60  
 MRNKEKSQRSQ +NSFKTIGIIGGGQLGQMMIAAIIYMGHKVITLDPASD PASRVSEVI  
 10 Sbjet: 1 MRNKEKSQRSQVNSFKTIGIIGGGQLGQMMIAAIIYMGHKVITLDPASDSPASRVSEVI 60

Query: 61 VAPYDDVEALGTLAARCDVLTVEFENVDADGLDAVVSAGQLPQGTDLRLISQNRIFEKDF 120  
 VAPYDDVEALG LAARCDVLTVEFENVDADGLDAVSA QLPGTDLRLISQNR I EKDF  
 15 Sbjet: 61 VAPYDDVEALGQLAARCDVLTVEFENVDADGLDAVVSACQLPQGTDLRLISQNRIVEKDF 120

Query: 121 LANKAGVTVPYKVVTSDDLGLDLTKTYVLKTATGGYDGHGQKVIRSAEDLPEAQQLA 180  
 LANKAGVTVPYKVVTSDDL GLDLTKTYVLKT TGGYDGHGQK+IRSAEDLPEAQQLA  
 Sbjet: 121 LANKAGVTVPYKVVTSDDLGLDLTKTYVLKTETGGYDGHGQKIIRSAEDLPEAQQLA 180

Query: 181 NSAQCVLEEFVNFDFLEISVIVSGNGQDVTVPVQENIHRNNILSKTIVPARISDQLADKA 240  
 NSAQCVLEEFVNFDFLEISVIVSGNG+DVTVPVQENIHRNNILSKTIVPARISDQLADKA  
 20 Sbjet: 181 NSAQCVLEEFVNFDFLEISVIVSGNGQDVTVPVQENIHRNNILSKTIVPARISDQLADKA 240

Query: 241 KEMAVQIAKKLQLSGTLCVEMFATADDIIVNEIAPRPHNSGHYSIEACDFSQFDTHILGV 300  
 K+ AVQIAKKLQLSGTLCVEMF TADDIIVNEIAPRPHNSG YSIEACDFSQFDTHILGV  
 25 Sbjet: 241 KKTAVQIAKKLQLSGTLCVEMFTTADDIIVNEIAPRPHNSGRYSIEACDFSQFDTHILGV 300

Query: 301 LGAPLPPIKLHAPAVMFNVLGQHVQQAIDHVAQNPSAHLHMYGKLEAKHNRMKGHVTVFS 360  
 LGAPLP I+LHAPAVM NVLGQHVQQA D+VA+NPSAHLHMYGKLEAKHNRMKGHVTVF+  
 30 Sbjet: 301 LGAPLPQIQLHAPAVMLNVLGQHVQQAIDYVAKNPSAHLHMYGKLEAKHNRMKGHVTVFA 360

Query: 361 DVPDEVEEF 369  
 DEV+EF  
 35 Sbjet: 361 KDADEVKEF 369

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 283

A DNA sequence (GBSx0310) was identified in *S.agalactiae* <SEQ ID 907> which encodes the amino acid  
 40 sequence <SEQ ID 908>. This protein is predicted to be phosphoribosylaminoimidazole carboxylase  
 catalytic subunit (purE). Analysis of this protein sequence reveals the following:

Possible site: 45  
 >>> Seems to have no N-terminal signal sequence

## 45 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3572(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

## 50 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12462 GB:Z99107 phosphoribosylaminoimidazole carboxylase I  
 [Bacillus subtilis]  
 Identities = 106/162 (65%), Positives = 128/162 (78%)

Query: 33 MQPIISIIMGSKSDWTTMQKTAEVLDNFGIAYEKKVVSARHTPDLMFKAHEEARGRGIKI 92  
 MQP++ IIMGS SDW TM+ ++LD + YEKKVVSARHTPD MF++AE AR RGIK+  
 55 Sbjet: 1 MQPLVGIIMGSTSDWETMKHACDILDELNVPEYKVVSAHRTPDFMFEYAETARERGIKV 60

Query: 93 IIAGAGGAHLPGMVAAKTTLPVIGVPVKSRLSGLDSLIVQMPGGVPVATMAIGEAG 152  
 IIAGAGGAHLPGM AAKTTLPVIGVPV+S+AL+G+DSL SIVQMPGGVPVAT +IG+AG  
 60

-358-

Sbjct: 61 IIAGAGGAAHLPGMTAAKTTLPVIGVPVQSKALNGMDSLISIVQMPGGVPVATTSIGKAG 120

Query: 153 ATNAALTALRILSIEDQNLADALAHFHEEQGKIAEESSNELI 194

A NA L A +ILS D++LA L E + ESS++L+

Sbjct: 121 AVNAGLLAAQILSAFDEDLARKLDERRENTKQTVLESSDQLV 162

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 909> which encodes the amino acid sequence <SEQ ID 910>. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -3.08 Transmembrane 36 - 52 ( 34 - 52)

----- Final Results -----

bacterial membrane --- Certainty=0.2232 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA04375 GB:AJ000883 purE [Lactococcus lactis]

Identities = 105/158 (66%), Positives = 131/158 (82%)

Query: 46 ISIIMGSKSDWATMQKTAEVLDNFGIAYEKKVSAHRTPDLMFKHABEARGGIKIIAG 105

++IIMG SDWATM++TA++LD+FG+AYEKKVSAHRT LM + + +AR RG K+IIAG

Sbjct: 4 VAIIMGCSDDWATMKETAKILDDFGIAYEKKVSAHRTPALMAEFSSQARERGYKVIIAG 63

Query: 106 AGGAAHLPGMVAAKTTLPVIGVPVKSRLSGLDSLISIVQMPGGVPVATMAIGEAGATNA 165

AGGAAHLPGMV+A+T +PVIGVP+KSRALSGLDSLISIVQMP GVPVATMAIGEAGA NA

Sbjct: 64 AGGAAHLPGMVSAQTLVPVIGVPIKSRLSGLDSLISIVQMPAGVPVATMAIGEAGAKNA 123

Query: 166 ALTALRILSIEDQNLADALAHFHEEQGKIAEESSGELI 203

AL AL++L+ ++NL L + ++ EES+ L+

Sbjct: 124 ALFALQLLANTNENLIQKLLVYRAAAQEMVEESNKALL 161

An alignment of the GAS and GBS proteins is shown below:

Identities = 162/169 (95%), Positives = 164/169 (96%), Gaps = 1/169 (0%)

Query: 27 PLYLNMQ-PIISIIMGSKSDWTTMQKTAEVLDNFGIAYEKKVSAHRTPDLMFKHAEEA 85

PL + IM+ PIISIIMGSKSDW TMQKTAEVLDNFGIAYEKKVSAHRTPDLMFKHAEEA

Sbjct: 35 PLCILIMKTPIISIIMGSKSDWATMQKTAEVLDNFGIAYEKKVSAHRTPDLMFKHAEEA 94

Query: 86 RGRGIKIIAGAGGAAHLPGMVAAKTTLPVIGVPVKSRLSGLDSLISIVQMPGGVPVAT 145

RGRGIKIIAGAGGAAHLPGMVAAKTTLPVIGVPVKSRLSGLDSLISIVQMPGGVPVAT

Sbjct: 95 RGRGIKIIAGAGGAAHLPGMVAAKTTLPVIGVPVKSRLSGLDSLISIVQMPGGVPVAT 154

Query: 146 MAIGEAGATNAALTALRILSIEDQNLADALAHFHEEQGKIAEESSNELI 194

MAIGEAGATNAALTALRILSIEDQNLADALAHFHEEQGKIAEESS ELI

Sbjct: 155 MAIGEAGATNAALTALRILSIEDQNLADALAHFHEEQGKIAEESSGELI 203

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 284

A DNA sequence (GBSx0311) was identified in *S.agalactiae* <SEQ ID 911> which encodes the amino acid sequence <SEQ ID 912>. This protein is predicted to be phosphoribosylglycinamide synthetase (purD). Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----



-359-

bacterial cytoplasm --- Certainty=0.1966(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

# 5 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA04374 GB:AJ000883 purD [Lactococcus lactis]

Identities = 236/419 (56%), Positives = 298/419 (70%), Gaps = 7/419 (1%)

10 Query: 1 MKLLVVGSGGREHAI AKKLLASKDQVVFVAPGNDGMTLDGLDLVNIGISEHSRLIDFVK 60  
 MK+LV+GSGGREHA+AKK + S V++VFVAPGN GM DG+ +V+I + +L+ F +  
 Sbjct: 1 MKILVIGSGGREHALAKKFME SPQVEEVFVAPGNSGMEKDGIQIVHISELSNDKLVKFAQ 60

15 Query: 61 ENEIAWTLIGPDDALAAGIVDGFNSAGLRAFGPTKAAAELEWSKDFAKEIMVKYNVPTAA 120  
 I T +GP+ AL G+VD F A L FGP K AAEE SKDFAK IM KY VPTA  
 Sbjct: 61 NQNIGLTFVGPETALMNGVVDFAFIKAE LPIFGPNKMAAELEGSKDFAK SIMKKYGVPTAD 120

20 Query: 121 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 180  
 Y TF E A AY++E+G P+V+KADGLA GKG V VA +E A A ++ F S  
 Sbjct: 121 YATFDSLEPALAYLDEKGVPLVIKADGLAAGKGVTVAFDIETAKSALADI-----FSGSQ 175

25 Query: 181 ARVVIEEFLDGEEFSLFAFANGDKFYIMPTAQDHKRAYDGDGKLGNTGGMGAYAPVPHLPQ 240  
 +VVIEEFLDGEEFSLF+F + K Y MP AQDHKRA+D DKG NTGGMGAY+PV H+ +  
 Sbjct: 176 GKVVIEEFLDGEEFSLFSFIHDGKIYPMPIAQDHKRAFDEDKGPNTGGMGAYSPVLHISK 235

30 Query: 241 SVVDTA VETIVKPVLEGMIAEGRPYLGVLVYAGLILTADGPKVIEFN SRFGDPETQIILPR 300  
 VV+ A+E +VKP + GMI EG+ + GVLVYAGLILT DG K IEFN+RFGDPETQ++LPR  
 Sbjct: 236 EVVNEALEKVVKPTVAGMIEEGKSFTGVLYAGLILTEDGVKTIEFNARFGDPETQVVLPR 295

35 Query: 301 LTSDF AQNIDIMMGIEPYITWQKDGVT LGVVVASEGYPLDYEKGVPLPEKTDGDIITYY 360  
 L SD AQ I DI+ G EP + W + GVT LGVVVA+EGYP + G+ LPE +G + YY  
 Sbjct: 296 LKSDLAQAIIDILAGNEPTLEWLESGVT LGVVVAAEGYPSQAKLGLILPEIPEG-LNVVY 354

Query: 361 AGAKFAENSKALLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQD TTGLFYRNDIGSKAI 419  
 AG EN++ L+S+GGRVY++ T + VK+ Q +Y +L + + G FYR+DIGS+AI  
 Sbjct: 355 AGVSKNENNQ-LISSGGRVYLVS ETGEDVKSTQKLLYEKLDKLENDGFFYRHDIGSRAI 412

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 913> which encodes the amino acid sequence <SEQ ID 914>. Analysis of this protein sequence reveals the following:

Possible site: 35

40 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.80 Transmembrane 5 - 21 ( 5 - 21)

----- Final Results -----

45 bacterial membrane --- Certainty=0.1319(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA04374 GB:AJ000883 purD [Lactococcus lactis]

Identities = 236/419 (56%), Positives = 301/419 (71%), Gaps = 7/419 (1%)

50 Query: 50 LKLLVVGSGGREHAI AKKLLASKGVDQVVFVAPGNDGMTLDGLDLVNIVVSEHSRLIAFAK 109  
 +K+LV+GSGGREHA+AKK + S V++VFVAPGN GM DG+ +V+I + +L+ FA+  
 Sbjct: 1 MKILVIGSGGREHALAKKFME SPQVEEVFVAPGNSGMEKDGIQIVHISELSNDKLVKFAQ 60

55 Query: 110 ENEISWAFIGPDDALAAGIVDDFN SAGLRAFGPTKAAAELEWSKDFAKEIMVKYNVPTAA 169  
 I F+GP+ AL G+VD F A L FGP K AAEE SKDFAK IM KY VPTA  
 Sbjct: 61 NQNIGLTFVGPETALMNGVVDFAFIKAE LPIFGPNKMAAELEGSKDFAK SIMKKYGVPTAD 120

60 Query: 170 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 229  
 Y TF E A AY++E+G P+V+KADGLA GKG V VA +E A A ++ F S  
 Sbjct: 121 YATFDSLEPALAYLDEKGVPLVIKADGLAAGKGVTVAFDIETAKSALADI-----FSGSQ 175

Query: 230 ARVVIEEFLDGEEFSLFAFANGDKFYIMPTAQDHKRAFDGDKGPNTGGMGAYAPVPHLPQ 289

-360-

+VVIEEFLDGEEFSLF+F + K Y MP AQDHKRAFD DKGPNITGGMGAY+PV H+ +  
 Sbjct: 176 GKVVIEEFLDGEEFSLFSGFIHDGKIYPMPIAQDHKRAFDEDDKGPNITGGMGAYSPVLHISK 235

Query: 290 SVVDTAVEMIVRPVLEGMVAEGRPYLGVLVGLILTADGPKVIEFNSRFGDPETQIILPR 349  
 VV+ A+E +V+P + GM+ EG+ + GVLV GLILT DG K IEFN+RFGDPETQ++LPR  
 Sbjct: 236 EVVNEALEKVVKPTVAGMIEEGKSFTGVLYAGLILTEDGVKTIEFNARFGDPETQVVLPR 295

Query: 350 LTSDFQAQNIDIMMGIEPYITWQKDGVTGLGVVASEGYPPDYEGVPLPEKTDGDIITYY 409  
 L SD AQ I DI+ G EP + W + GVTGLGVVVA+EGYP + G+ LPE +G + YY  
 Sbjct: 296 LKSDLAQAIIDILAGNEPTLEWLESQVTLGVVVAEGYPSQAKLGLILPEIPEG-LNVYY 354

Query: 410 AGVKFSENSELLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQDITGLFYRNDIGSKAI 468  
 AGV +EN++ L+S+GGRVY++ T + VK+ Q +Y +L + + G FYR+DIGS+AI  
 Sbjct: 355 AGVSKNENNQ-LISSGGRVYLVSSETGEDVKSTQKLLYKLDKLENDGFFFYRHDIGSRAI 412

An alignment of the GAS and GBS proteins is shown below:

Identities = 399/421 (94%), Positives = 408/421 (96%)

Query: 1 MKLLVVGSGGREHAIKLLASKDQVQVFPAGNDGMTLDGLDLVNIGISEHSRLIDFVK 60  
 +KLLVVGSGGREHAIKLLASK VDQVFPAGNDGMTLDGLDLVNI +SEHSRLI F K  
 Sbjct: 50 LKLLVVGSGGREHAIKLLASKQVQVFPAGNDGMTLDGLDLVNIIVVSEHSRLIAFAK 109

Query: 61 ENEIAWTLIGPDDALAAGIVDFNSAGLRAFQPTKAAAELEWSKDFAKEIMVKYNVPTAA 120  
 ENEI+W IGPDDALAAGIVD FNSAGLRAFQPTKAAAELEWSKDFAKEIMVKYNVPTAA  
 Sbjct: 110 ENEISWAFIGPDDALAAGIVDDFNSAGLRAFQPTKAAAELEWSKDFAKEIMVKYNVPTAA 169

Query: 121 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 180  
 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG  
 Sbjct: 170 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 229

Query: 181 ARVVIEEFLDGEEFSLFAFANGDKFYIMPTAQDHKRAYDGDGKGLNTGGMGAYAPVPHLPQ 240  
 ARVVIEEFLDGEEFSLFAFANGDKFYIMPTAQDHKRA+DGDGK NTGGMGAYAPVPHLPQ  
 Sbjct: 230 ARVVIEEFLDGEEFSLFAFANGDKFYIMPTAQDHKRAFDGDKPNTGGMGAYAPVPHLPQ 289

Query: 241 SVVDTAVETIVKPVLEGMIAEGRPYLGVLVYAGLILTADGPKVIEFNSRFGDPETQIILPR 300  
 SVVDTAVE IV+PVLEGM+AEGRPYLGVLV GLILTADGPKVIEFNSRFGDPETQIILPR  
 Sbjct: 290 SVVDTAVEMIVRPVLEGMVAEGRPYLGVLVYAGLILTADGPKVIEFNSRFGDPETQIILPR 349

Query: 301 LTSDFQAQNIDIMMGIEPYITWQKDGVTGLGVVASEGYPLDYEGVPLPEKTDGDIITYY 360  
 LTSDFQAQNIDIMMGIEPYITWQKDGVTGLGVVASEGYP DYKGVPLPEKTDGDIITYY  
 Sbjct: 350 LTSDFQAQNIDIMMGIEPYITWQKDGVTGLGVVASEGYPPDYEGVPLPEKTDGDIITYY 409

Query: 361 AGAKFAENSKALLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQDITGLFYRNDIGSKAIKE 421  
 AG KF+ENS+ LLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQDITGLFYRNDIGSKAI+E  
 Sbjct: 410 AGVKFSENSELLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQDITGLFYRNDIGSKAIRE 470

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 285

A DNA sequence (GBSx0312) was identified in *S.agalactiae* <SEQ ID 915> which encodes the amino acid sequence <SEQ ID 916>. Analysis of this protein sequence reveals the following:

Possible site: 36  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.28 Transmembrane 235 - 251 ( 235 - 251)

----- Final Results -----  
 bacterial membrane --- Certainty=0.1510(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

-361-

&gt;GP:AAA23257 GB:M81878 unknown [Clostridium perfringens]

Identities = 66/258 (25%), Positives = 119/258 (45%), Gaps = 9/258 (3%)

Query: 1 MTIYDQIESALDMLTDLEREIACYFMGQPIISKDALASTIVTKQLHISQAALTRFAKKCGF 60  
 M I +Q+E+ T E+ + Y + + +I+ K+ +A +TRF KK GF  
 Sbjct: 1 MGILEQLENPKFKATKSEKTLIEYIKSDLDNIIYKSISIIAKESGVGEATITRFTKKLGF 60

Query: 61 KGYREFVFEYLKS-HETISQQLYGLQNDNTKKVFMNYQEMISKSADI-----IDEEQL 112  
 G+++F K + + L + V +M+ S +I ID + +  
 Sbjct: 61 NGFQDFKVTLAKEISNKKNTSIINLHVHRDESVTETANKMLKSSINILEQTVKQIDLDLM 120

Query: 113 LEVSHMIEQADRVYFYGKGSSSLVAKEFKIRLMRLGVICEALDDTDSFSWINSIVNDRCL 172  
 + +I A RVYF G G S + A + + MR+G + D+ + +SI ND +  
 Sbjct: 121 CKCRDLIMNAKRVIYFIGYSGIAATDINXKFMRIQFTTVPVTDSTMTVMSSITNDDDV 180

Query: 173 VIAFSLSGNTNSVIGALKIASCHGAKTVLFTK-QPHTIDYAFDKIIQVASARHLDYGNRI 231  
 ++A S SG T VI +K A +G K + T+ + + D + SA + I  
 Sbjct: 181 IVAISNSGTTKEVIKTVKQAKENGTKIITLTEDSDNPLRLKSDYELTYTSAETIFETGSI 240

Query: 232 SPQIPMLIMVDIIYAQFL 249  
 S +IP + ++D++Y + +  
 Sbjct: 241 SSKIPQIFLLDLLYTEVI 258

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 917> which encodes the amino acid sequence <SEQ ID 918>. Analysis of this protein sequence reveals the following:

Possible site: 60

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -4.88 Transmembrane 243 - 259 ( 242 - 261)

----- Final Results -----

bacterial membrane --- Certainty=0.2954(Affirmative) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial cytoplasm --- Certainty=0.0000(Not Clear) &lt; succ&gt;

A related sequence was also identified <SEQ ID 9093> which encodes the amino acid sequence <SEQ ID 9094>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 56

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -4.88 Transmembrane 239 - 255 ( 238 - 257)

----- Final Results -----

bacterial membrane --- Certainty= 0.295(Affirmative) &lt; succ&gt;

bacterial outside --- Certainty= 0.000(Not Clear) &lt; succ&gt;

bacterial cytoplasm --- Certainty= 0.000(Not Clear) &lt; succ&gt;

An alignment of the GAS and GBS proteins is shown below:

Identities = 138/263 (52%), Positives = 189/263 (71%), Gaps = 2/263 (0%)

Query: 6 QIESALDMLTDLEREIACYFMGQPIISKDALASTIVTKQLHISQAALTRFAKKCGFKGYRE 65  
 +IE++L+ MT LE+ IA +F+ ++ L ++ + K+LHISQAALTRFAKKCGF GYR  
 Sbjct: 14 KIEASLEHMTSLEKGAHFFITTDLTPOELTASEIVKRLHISQAALTRFAKKCGFTGYRA 73

Query: 66 FVFEYLKSHETISQQLYGLQNDNTKKVFMNYQEMISKSADIIDEEQLLEVSHMIEQADRV 125  
 F F+YL S + + + TK+V M+Y +I+K+ ++++EE+LL ++ +I+ ++RV  
 Sbjct: 74 FAFDYHLSQLQESQETFSIHLELTTRVLMYDALINKTYELVNEEKLLNLAKLIDSSERV 133

Query: 126 YFYGKGSSSLVAKEFKIRLMRLGVICEALDDTDSFSWINSIVNDRCLVIAFSLSGNTNSV 185  
 YF+GKGSS LVA+E K+R MRLG+IC+A DTD F+W NS+VN+ CLV FSLSG TNSV  
 Sbjct: 134 YFFGKGSSGLVAREMKLRFMRGLGICDAYSDDTGFTWANSVLNENCLVFGFSLSGKNTNSV 193

Query: 186 IGALKIASCHGAKTVLFTKQPHT-IDYAFDKIIQVASARHLDYGNRISPQIPMLIMVDII 244  
 I AL AS GAKTVL T T D + D II V+S L YGNR+SPQ P+LIM+DII  
 Sbjct: 194 ITALHQASQRGAKTVLLTTDNQTEFDDSLD-IIPVSSTHQLHYGNRVSPQFPLLIMMDII 252

-362-

Query: 245 YAQFLDINKIEKERIFRETTIQR 267  
 YA L I+K KE+IF+ TII +  
 Sbjct: 253 YAYVLAIDKPHKEKIFKNTIIDK 275

SEQ ID 916 (GBS320) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 40 (lane 5; MW 33kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 7; MW 58kDa) and in Figure 160 (lane 7 & 8; MW 58kDa).

GBS320-GST was purified as shown in Figure 224, lane 3-4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 286

A DNA sequence (GBSx0313) was identified in *S.agalactiae* <SEQ ID 919> which encodes the amino acid sequence <SEQ ID 920>. This protein is predicted to be xylan esterase 1 (cephalosporin-C). Analysis of this protein sequence reveals the following:

Possible site: 54  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4981(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB68821 GB:AF001926 xylan esterase 1 [Thermoanaerobacterium sp.  
 'JW/SL YS485']  
 Identities = 133/299 (44%), Positives = 188/299 (62%), Gaps = 1/299 (0%)

Query: 5 MSLDDMREYLGQDQIPEDFDFFWKQTMKYQG-NIEYRLDKKDFNITFAQAYDLHFKGSN 63  
 M L +REY G + PEDFD++W + + + + L + F ++FA+ YDL+F G  
 Sbjct: 6 MPLQLREYTGTPCEDFDEYWNRAIDEMRSVDPKIELKESSFQVSFAECYDLYFTGVR 65

Query: 64 NSIVYAKCLFPKTNKPYPVVFYFHGYQNQSPDWSQDQINYVAAGYGVVSMVVRGQAGQSQD 123  
 + ++AK + PKT +P + FHGY + S DW+D+LNYVAAG+ VV+MDVRGQ GQSQD  
 Sbjct: 66 GARIHAKYIKPKTEGKHPALIRFHGYSSNSGDWNDKLNVAAGFTVVAMVVRGQGGSQD 125

Query: 124 KGHFDGITVKGQIVRGMISGPNHLFYKDIYLDVDFQLIDIIATLESVDNQLYSYGWSQGG 183  
 G G T+ G I+RG+ +++ ++ I+LD QL I+ + VD +++ G SQGG  
 Sbjct: 126 VGGVTGNTLNGHIIRGLDDADNMLFRHIFLDTAQLAGIVNMPEVDEDRVGVMGPSQGG 185

Query: 184 ALALIAAALNPKIVKTAVYPFLSDFRRVLDLGGVSEPYDELFRYFKYSDFPHKTENNVL 243  
 L+L AAL P++ K V+ YPFLSD++RV DL Y E+ YF+ DP H+ EN V  
 Sbjct: 186 GLSLACAALEPRVRKVSEYPFLSDYKRVWDLDLAKNAYQEITDYFRLFDPRHERENEVF 245

Query: 244 KTLAYIDVKNFAHRISCPVLLTALKDDICPPSTQFAIFNRLTSTKKHLLLPDYGHDP 302  
 L YIDVKN A RI V++ L D +CPPST FA +N + S K + PDYGH+PM  
 Sbjct: 246 TKLGIDVKNLAKRIKGDVLMCVLMDQVCPSTVFAAYNNIQSKKDIKVPDYGHDP 304

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-363-

**Example 287**

A DNA sequence (GBSx0314) was identified in *S.agalactiae* <SEQ ID 921> which encodes the amino acid sequence <SEQ ID 922>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -5.73    Transmembrane 128 - 144 ( 126 - 145)

----- Final Results -----
                bacterial membrane --- Certainty=0.3293(Affirmative) < succ>
10                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15 >GP:AAA23256 GB:M81878 unknown [Clostridium perfringens]
    Identities = 78/160 (48%), Positives = 110/160 (68%)

Query: 131 CLTIGTGIGGCLIIDKTVFHGFSNSACEVGYMHLSDGDFQDLASTTALIADVAKAHGDEI 190
          CLTIGTGIGG LIID V HGFSNSA E+GYM ++ + QD+AS +AL+ +VA G E
Sbjct: 18 CLTIGTGIGGALIIDGKVLHGFSNSAGEIGYMMVNGENIQDIASASALVKNVALRKGVEP 77

20 Query: 191 SRWDGRRIFQEAKKGNEKCIASIDRMINYLGQGIANMVYVNPKEVVLGGGIMAQKDYLO 250
          S DGR + + G+ C +++++ + L GI+N+VY++NPE VVLGGGIMA+++ +
Sbjct: 78 SSIDGRYVLDNYENDLICKEEVEKLADNLALGISNIVYLINPEVVVLGGGIMAREEVFR 137

25 Query: 251 DKLSESLKRNLTSLAEKTAIVFAQHENQAGMLGAYYHFK 290
          + SL++ L+ S+ T I FA+ +N AGM GAYY+FK
Sbjct: 138 PLIENSLRKYLIESVYNNTKIAFAKLKNTAGMKGAYYNFK 177

```

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 923> which encodes the amino acid sequence <SEQ ID 924>. Analysis of this protein sequence reveals the following:

```

Possible site: 22

>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -4.30    Transmembrane 128 - 144 ( 127 - 145)
35    INTEGRAL    Likelihood = -0.11    Transmembrane 227 - 243 ( 227 - 243)

----- Final Results -----
                bacterial membrane --- Certainty=0.2720(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
40                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

45 >GP:BAB04516 GB:AP001509 glucose kinase [Bacillus halodurans]
    Identities = 97/291 (33%), Positives = 155/291 (52%), Gaps = 14/291 (4%)

Query: 5 LAIDIGGTAIKYGLISETGDLLKEEMATEAYKGGPSILEKVKGLVKTYQDQMDLAGVAI 64
          + ID+GGT IK L+S+ G+++ +E TEA +G ++ K+ L + D AG+ I
Sbjct: 3 VGIDLGGTIKIAALVSDAGEIISVQECPTAAQGPPEVMNKMMSLTEKVTDHQPFAGIGI 62

50 Query: 65 SSAGMVNPDEGEIFYAGQPINPYAGTQFKKEIETTFGLPCEVENDVNCAGLAEAISGSAK 124
          + G ++ EG I + P +P + +E F P +++ND N A LAEA+ GS +
Sbjct: 63 GAPGPLSSTEGTIL-SPPNLPGWDDHILVDRFQEQFQCPVKLDNDANVAALAEALLGSGQ 121

55 Query: 125 DYPVALCLTIGTGIGGCLLFNSQVPHGSSHSACEVG-----YLHLSDGQFQDLAS 174
          + LTI TGIGG + + + HG+S A E+G + +L+ G + LAS
Sbjct: 122 GFTSVFYLTISTGIGGGYVLDGSIHVHGASDYAGEIGNMIVQPNGYQHANLNPGLSLEGLAS 181

Query: 175 TTALVQVEVVLAYGDDISQWDGRRIFEQAKAGDAICIAAISQVDYLGQGIANICYVVPNP 234
          TA+ + +G + R +F+Q + GD + + +DYL GIANI + +NP+
60 Sbjct: 182 GTAIGRMARERFG---VEGGTREVFDQIRRGDHDQMRLVEEAMDYLAIGIANIAHTINPD 238

```

-364-

Query: 235 VVVLGGGIMAQKDYLDKLTALDSYLVSSLAKKTQLKFASHGNNAGILGA 285  
 V VLGGG+M D + +K + YL LA+ T + A G ++G+LGA  
 Sbjct: 239 VFVLGGGVMMADDLILPIVKEKVSRYLYPGLAQSTTIVKAKLGGDSGVILGA 289

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 192/292 (65%), Positives = 237/292 (80%)

Query: 1 MTRTVAIDIGGTMIKHGIVDNLGCIVEASELATEAYKGGPGILQKVCQIIDNYLAEGSID 60  
 M +AIDIGGT IK+G++ G ++E E+ATEAYKGGP IL+KV ++ Y + +  
 10 Sbjct: 1 MKHYLAIDIGGTAIKYGLISETGDLLEKEEMATEAYKGGPSILEKVKGLVKTYQDQMDLA 60

Query: 61 GIAISSAGMVPDEGCIFYSGPQIPNYAGTQFKKVLDTYQVTEIENDVNCAGLAEAVS 120  
 G+AISSAGMV+PDEG IFY+GPQIPNYAGTQFKK +E+T+ + E+ENDVNCAGLAEA+S  
 15 Sbjct: 61 GVAISSAGMVNPDEGEIFYAGPQIPNYAGTQFKKEIETFGLPCEVENDVNCAGLAEAIS 120

Query: 121 GSAKDSSIALCLTIGTGIGGCLIIDKTIVFHGFSNSACEVGYMHLSDGDFQDLASTTALIA 180  
 GSAKD +ALCLTIGTGIGGCL+ + VFHG S+SACEVGY+HLSDG FQDLASTTAL+  
 20 Sbjct: 121 GSAKDYPVALCLTIGTGIGGCLLFNSQVFGSSHSACEVGYLHLSGQFQDLASTTALVQ 180

Query: 181 DVAKAHGDEISRWDGRRIFQEAKKGNEKCIASIDRMINYLGQGIANMVYVNPKEKVVVLGG 240  
 +V A+GD+IS+WDGRRIF++AK G+ CIA+I + ++YLQGIAN+ YVNP VVLGG  
 25 Sbjct: 181 EVVLAYGDDISQWDGRRIFEQAKAGDAICIAAISQVDYLGQGIANICYVNPVNVVLGG 240

Query: 241 GIMAQKDYLDKLSLKRNLVTSIAEKTAIVFAQHENQAGMLGAYYHFKNR 292  
 GIMAQKDYLDKL +L LV+SLA+KT + FA H N AG+LGAYYHFK +  
 25 Sbjct: 241 GIMAQKDYLDKLTALDSYLVSSLAKKTQLKFASHGNNAGILGAYYHFKQK 292

SEQ ID 922 (GBS331) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell  
 extract is shown in Figure 60 (lane 2; MW 35.9kDa). It was also expressed in *E.coli* as a GST-fusion  
 30 product. SDS-PAGE analysis of total cell extract is shown in Figure 67 (lane 3; MW 61kDa).

The GBS331-GST fusion product was purified (Figure 209, lane 3) and used to immunise mice. The  
 resulting antiserum was used for FACS (Figure 309), which confirmed that the protein is immunoaccessible  
 on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 35 vaccines or diagnostics.

### Example 288

A DNA sequence (GBSx0315) was identified in *S.agalactiae* <SEQ ID 925> which encodes the amino acid  
 sequence <SEQ ID 926>. This protein is predicted to be a acylneuraminate lyase (nanA). Analysis of this  
 protein sequence reveals the following:

40 Possible site: 18  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 45 bacterial cytoplasm --- Certainty=0.0894 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

50 >GP:CAA69950 GB:Y08695 putative acylneuraminate lyase [Clostridium  
 tertium]  
 Identities = 162/225 (72%), Positives = 191/225 (84%)

Query: 1 MKDLQKYQGIIPAFYACYDDKGDICPERVKALTNFYIDKGVQGLYVNGSSGECIYQSVAD 60  
 M++L+KY+GIIPAFYACYDD+G I PER + T Y IDKGV+GLYV GSSGECIYQS +  
 55 Sbjct: 1 MRNLEKYKGIIPAFYACYDDEGKISPRTQMFTQYLIDKGVKGLYVCGSSGECIYQSKEE 60

-365-

Query: 61 RKLVLENVMSVAKGKLTIVIAHVACNNTKDSVELAMHAEAGVDAIAAIPPIYFRLPEYAI 120  
 RK+ LENVM VAKGK+T+IAHV CNNT+DS ELA HAE+IGVDAIA+IPPIYF LP+Y+I  
 Sbjct: 61 RKITLENVMKVAKGKITIIAHVGCNNTRDSEELAEHAEISIGVDAIASIPPIYFHLPDYSI 120

Query: 121 ADYWNTISQAAPQTDFTIYNIPQLAGVALTSDLYRKMLQNPQVIGVKNSSMPVQDIQNFV 180  
 A+YWN IS AAP TDFIYNIPQLAGV L +LY++ML+NP+VIGVKNSSMPVQDIQ F  
 Sbjct: 121 AEYWNDISNAAPNTDFTIYNIPQLAGVGLGINLYKQMLKNPRVIGVKNSSMPVQDIQMFK 180

Query: 181 AIGGENHIVFNGPDEQFLGGRLMGAAAGIGGTYGVMPPELYLTNLQ 225  
 I G+ +VFNGPDEQF+ GR+MGA GIGGTY VMPEL+L ++  
 Sbjct: 181 DISGDES VVFNPGPDEQFVAGRIMGADGGIGGTYAVMPELFLAADK 225

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 927> which encodes the amino acid  
 sequence <SEQ ID 928>. Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0981(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 238/304 (78%), Positives = 263/304 (86%)

Query: 1 MKDLQKYQGIIPAFYACYDDKGDICPERVKALTNFYIDKGVQGLYVNGSSGECIYQSVAD 60  
 M DL KYQGIIPAFYACYDD+G+I PERV+ALT Y+IDKGVQGLY+NGSSGECIYQSV D  
 Sbjct: 1 MTDLTQYQGIIPAFYACYDDQGNISPERVRALTQYYIDKGVQGLYINGSSGECIYQSVFD 60

Query: 61 RKLVLENVMSVAKGKLTIVIAHVACNNTKDSVELAMHAEAGVDAIAAIPPIYFRLPEYAI 120  
 R+LVLENVM+VAKGKLT+I HVACNNTKDS+ELA H+E +GVDAIAAIPPIYFRLPEYA+  
 Sbjct: 61 RQLVLENVMAVAKGKLTIIHVACNNTKDSIELAAHSERLGVDAIAAIPPIYFRLPEYAV 120

Query: 121 ADYWNTISQAAPQTDFTIYNIPQLAGVALTSDLYRKMLQNPQVIGVKNSSMPVQDIQNFV 180  
 ADYWN IS AAP TDFIYNIPQLAGVALT LY+ ML N +VIGVKNSSMPVQDIQ F  
 Sbjct: 121 ADYWNAISSAAPHDTFTIYNIPQLAGVALTSPSLYKTM LANKRVIGVKNSSMPVQDIQTFC 180

Query: 181 AIGGENHIVFNGPDEQFLGGRLMGAAAGIGGTYGVMPPELYLTNLQLIVDKDLEKARELQF 240  
 AIGG++HIVFNGPDEQFLGGRLMGAAAGIGGTYG MPPEL+L LNQLI DKDLEKA+ LQ+  
 Sbjct: 181 AIGGDDHIVFNGPDEQFLGGRLMGAAAGIGGTYGAMPELFLRLNQLIADKDLEKAKALQY 240

Query: 241 TINDIITKLCSGHGNMYAVIKAVLEINEQLTIGSVRLPLASVTEEDKPIIKEAAEMIRHA 300  
 TIN+II L S HGNMY VIK VL INE L IGSVR PLA + EED+ I + AA +I A  
 Sbjct: 241 TINEIIGVLVSAHGNMYGVIKEVLRINEGLDIGSVRSPLAELVEEDRVICQRAAALINQA 300

Query: 301 KKQF 304  
 K+ F  
 Sbjct: 301 KETF 304

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 289

A DNA sequence (GBSx0317) was identified in *S.agalactiae* <SEQ ID 929> which encodes the amino acid  
 sequence <SEQ ID 930>. Analysis of this protein sequence reveals the following:

Possible site: 46  
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -9.45	Transmembrane	82 - 98 ( 79 - 111)
INTEGRAL	Likelihood = -6.85	Transmembrane	24 - 40 ( 21 - 52)
INTEGRAL	Likelihood = -5.26	Transmembrane	180 - 196 ( 172 - 200)

-366-

INTEGRAL Likelihood = -5.10 Transmembrane 160 - 176 ( 158 - 179)  
 INTEGRAL Likelihood = -4.35 Transmembrane 110 - 126 ( 106 - 130)

## ----- Final Results -----

5 bacterial membrane --- Certainty=0.4779(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:BA05827 GB:AP001514 unknown conserved protein in B. subtilis  
 [Bacillus halodurans]  
 Identities = 40/148 (27%), Positives = 74/148 (49%), Gaps = 4/148 (2%)

15 Query: 14 VNNPFMQGCNVVFDLALLNLLFMI-TCLPLVTIG--AAKISLYRTLWQKLEGD-QTNLLI 69  
 +++ F Q C+ ++ LA +NLL++ T L LV +G A +++ L + G+ +  
 Sbjet: 6 MSSRFYQTCDWIWKLAYINLLWLSGTLGLVVLGFLPATTAMFTVLRKWFNGPDVAITR 65

Query: 70 LYIKHLKKEWFQGMMLLGLVELSILVVIIFDLTILHYQIGFIVSFLKITCYAFLLLTVMTS 129  
 + + K E+ + LLG V L ++ F+ L G + L + YAFL+L ++T  
 20 Sbjet: 66 TFFQAYKNEFLKINLLGAVLLLGAYILYFNMYLGTVEGTVMVLSLGGWYAFLLIYIITL 125

Query: 130 IYLFPPMAARYEMSLDITVKKSFIMACLN 157  
 Y+ P Y + L +K + I+ +N  
 25 Sbjet: 126 FYIIPAYVHYNLKLFFQYIKTALIIGFVN 153

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 931> which encodes the amino acid sequence <SEQ ID 932>. Analysis of this protein sequence reveals the following:

## Possible site: 24

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

30 INTEGRAL Likelihood = -14.86 Transmembrane 117 - 133 ( 108 - 139)  
 INTEGRAL Likelihood = -7.48 Transmembrane 30 - 46 ( 21 - 54)  
 INTEGRAL Likelihood = -6.90 Transmembrane 88 - 104 ( 83 - 105)  
 INTEGRAL Likelihood = -6.26 Transmembrane 165 - 181 ( 151 - 187)  
 35 INTEGRAL Likelihood = -5.89 Transmembrane 189 - 205 ( 182 - 207)

## ----- Final Results -----

40 bacterial membrane --- Certainty=0.6944(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BA05582 GB:AP001513 unknown conserved protein in bacilli  
 [Bacillus halodurans]  
 Identities = 59/194 (30%), Positives = 93/194 (47%), Gaps = 11/194 (5%)

45 Query: 17 SKWMRASAAFLDLLVFNLLFVL-SCLPLLTIGV--AKMALYASLLDWREGQVS-QLVTTY 72  
 +K M+ + L+ NLL++L S + + +GV A +L+A W + + L TY  
 Sbjet: 8 TKIMKLFWEIMRLVYLNLLWLLFSFIGIILGVMPATASLFAVFRKWKYQKEDDFPLEQTY 67

50 Query: 73 SSHFKYYFKSGLRLGLIELGIMTICLLDLFLIRNQSGLVFQGFVKVLCVAVLFLVILFLY 132  
 + FK FK +GL + I I LD+ L+ S + Q + A+ F+ ++ LY  
 Sbjet: 68 LNEFKRSFKIANLVGLTIVLIGGILYLDVLLLLGTSHWIGQLLLMGVGALSFIYLVTLTY 127

55 Query: 133 AYPQAVKRDLSLSTLFKRSFLLAGLFFPWSFAFLAFLCTIFSLQL----SLLTLFGGVS 188  
 +P V DLS FK SFL L G+ P+ L I L++ +L LL LF S  
 Sbjet: 128 IFPTLVHFDLSYQYFKHSFLL-GVLQPF--TLLMITLSLSALLFLTFPILLPLF-AAS 184

Query: 189 LLAIIGISSLTLY 202  
 +A + + S + Y  
 60 Sbjet: 185 FMAALTMWSFLFGY 198

An alignment of the GAS and GBS proteins is shown below:

Identities = 68/210 (32%), Positives = 117/210 (55%)



-367-

Query: 3 KANQLIAAIFDVNNPFMQGCNVVFDLALLNLLFMITCLPLVTIGAAKISLYRTLWQKLEG 62  
 K L+ ++F +++ +M+ +FDL + NLLF+++CLPL+TIG AK++LY +L EG  
 Sbjct: 4 KKQGLLHSLFKLDSKWMRASAAFLDLLVFNNLLFVLSCLPLLTIGVAKMALYASLLDWREG 63

Query: 63 DQTNLLILYIKHLKKEWFQGMLLGLVELSILVVIIFDLTILHYQIGFIVSFLKITCYAFL 122  
 + L+ Y H K + G+ LGL+EL I+ + + DL ++ Q G + K+ C A L  
 Sbjct: 64 QVSQVLVTYSSHFYFKSGLRLGLIELGIMTICLLDLFLIRNQSGLVFQGFVKVLCVAVL 123

Query: 123 LLTVMTSIYLFPMARYEMSLDDTVKKSFIMACNLKWTGVLMLLIMTWFMVQSSLLF 182  
 L V+ +Y +P A + ++SL K+SF++A L W+ + + +T F + S L  
 Sbjct: 124 FLVVILFLYAYPQAVKRDLSLSTLFKRSFLLAGLFFPWSFAFLAFICLTIFSLQLSLLTL 183

Query: 183 MLTVSAIFIFAYTAFAYFKIIILQKQFAYF 212  
 VS + I ++ Y +II++ F  
 Sbjct: 184 FGGVSLLAIIIGISSLTLYLYLIIMESLLRRF 213

A related GBS gene <SEQ ID 8535> and protein <SEQ ID 8536> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 2  
 McG: Discrim Score: 3.27  
 GvH: Signal Score (-7.5): -4.23  
 Possible site: 46  
 >>> Seems to have an uncleavable N-term signal seq

ALOM program count: 5 value: -9.45 threshold: 0.0

INTEGRAL	Likelihood = -9.45	Transmembrane	82 - 98 ( 79 - 111)
INTEGRAL	Likelihood = -6.85	Transmembrane	24 - 40 ( 21 - 52)
INTEGRAL	Likelihood = -5.26	Transmembrane	180 - 196 ( 172 - 200)
INTEGRAL	Likelihood = -5.10	Transmembrane	160 - 176 ( 158 - 179)
INTEGRAL	Likelihood = -4.35	Transmembrane	110 - 126 ( 106 - 130)
PERIPHERAL	Likelihood = 5.89		142

modified ALOM score: 2.39

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial membrane	---	Certainty=0.4779(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the databases:

ORF00072(364 - 828 of 1260)  
 EGAD|108353|BS3003(14 - 171 of 222) hypothetical protein {Bacillus subtilis}  
 OMNI|NT01BS3507 conserved hypothetical protein GP|2635493|emb|CAB14987.1||Z99119 similar to  
 hypothetical proteins from B. subtilis {Bacillus subtilis}  
 GP|2293197|gb|AAC00275.1||AF008220 YteU {Bacillus subtilis} PIR|D69991|D69991 conserved  
 hypothetical protein yteU - Bacillus subtilis  
 %Match = 5.9  
 %Identity = 26.6 %Similarity = 50.6  
 Matches = 42 Mismatches = 74 Conservative Sub.s = 38

270	300	330	360	390	417	441	471
IMS	KKG	Y*KC	WRKKYREYIVK	KANQLIAAIFDVNNPFMQGCNVVFDLALLNLLFMI	-TCLPLVTIG--AAKISLYRTLW		
					: :       : :         : : :		
					MEHDGSLGRMLRFCEWIMRFAYTNLLWLFFTLGLGVFGIMPATAALFAVMR		
			10	20	30	40	50

498	528	558	588	618	648	678	708
QKLEG-DQTNLLILYIKHLKKEWFQGMLLGLVELSILVVIIFDLTILHYQIGFIVSFLKITCYAFL	LLTVMTSIYLFPM						
:	:	:	:	:	:	:	:
KWIQGD	NVPVLK	TFWQ	EYKGEFFRSNLLGAVLALIGV	IYIDLALI-YPSHFLHLIRFAIMIFGFLFVSM	LFYVFP	LL	
	70	80	90	100	110	120	130

738	768	798	828	858	888	918	948
ARYEMSLDDTVKKS	FIMACNLKWTGVL	MLLIMTWFMVQSSLLF	MLTVSAIFIFAYTAFAYFKIIILQKQFAYFSKQ				

-368-

```

::      || |:::  |::|  :: | :  :|::
VHFDWKKRLYVKFSLLSVAYLQYTLTMLALTVALFFLLAYLPGIVPFFSVSLISYCHMRIVYAVLLKVEQHGGEPRKS
          150          160          170          180          190          200          210

```

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 290

A DNA sequence (GBSx0318) was identified in *S.agalactiae* <SEQ ID 933> which encodes the amino acid sequence <SEQ ID 934>. Analysis of this protein sequence reveals the following:

```

10 Possible site: 51
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.1827(Affirmative) < succ>
15    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

20 >GP:AAC44392 GB:U43526 ORF-1 [Streptococcus pneumoniae]
    Identities = 48/151 (31%), Positives = 66/151 (42%), Gaps = 5/151 (3%)

    Query: 1  MIYDHLNLTHYKDINPNLDAIDYLLSHDLRNLDIGTYHISPEVILMVQSNQLSES-FD 59
             MI  +  L  Y  +NP+  ID+L  L  NL  G+  I  +  L++
    Sbjct: 1  MIITKISRLGTYVGVNPHFATLIDFLEKTGLENLTEGSIADGNRLFGNCFTYLADGQAG 60

25    Query: 60  HIFEYHKYLDIHVIEGHEVIKLGKGDKEV-EEY--LGDIGFIKSEETSFDLRDNYI 116
             FE  H+KYLDIH  V+E  E  +  +  +  V  V  +EY  DI  E  LR
    Sbjct: 61  AFFETHQKYLDIHLVLENEEAMAVTSPENVSVTQEYDEEKDIELYTGKVEQLVHLRAGEC 120

30    Query: 117  AFFFPEEAHQPNMGSLGNYVKKGVKVLKVLMA 147
             FPE+ HQP  +  VKK  V  KV  ++
    Sbjct: 121  LITFPEDLHQPK-VRINDEPVKKVVFKVAIS 150

```

No corresponding DNA sequence was identified in *S.pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 291

A DNA sequence (GBSx0319) was identified in *S.agalactiae* <SEQ ID 935> which encodes the amino acid sequence <SEQ ID 936>. This protein is predicted to be sugar ABC transporter, permease protein (araQ).

- 40 Analysis of this protein sequence reveals the following:

```

    Possible site: 35
    >>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -7.38    Transmembrane  245 - 261 ( 239 - 265)
    INTEGRAL    Likelihood = -3.72    Transmembrane  140 - 156 ( 139 - 158)
45    INTEGRAL    Likelihood = -3.61    Transmembrane   76 -  92 (  71 -  94)
    INTEGRAL    Likelihood = -2.81    Transmembrane  112 - 128 ( 107 - 128)
    INTEGRAL    Likelihood = -1.59    Transmembrane  188 - 204 ( 186 - 204)

    ----- Final Results -----
50    bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

-369-

>GP:AAD35515 GB:AE001721 sugar ABC transporter, permease protein  
[Thermotoga maritima]

Identities = 94/262 (35%), Positives = 158/262 (59%), Gaps = 1/262 (0%)

5 Query: 15 LILCLLTVLFIFFPYWIMTGAFKSQPDIIIPPQWWPKAPTLENFKALTVQNPAIRLWN 74  
+ + + V+F+ P ++ + +FK + PP +PK P+LE + + + L+L N  
Sbjct: 9 IFIVFMLVVFMLPVFYAVVSSFKPMSEIYSYPPTIFPKKPSLEGYINVIKEYDLLTYLRN 68

10 Query: 75 SVFISIMTMFLVCCTSSMAGYVLAKKRFYQKILFSLFIAAMALPKQVVLVPLVRIINFM 134  
++F++ + + S M GY LAK +F+G + + S+F M + QV++VPL +I +  
Sbjct: 69 TLFVATVATVITVLVSVMTGYGLAKGKFWGIRPVNSMFTMTMFVSAQVIMVPLFVVIRSL 128

15 Query: 135 GIHDTLWAVILPLVGWPFVFLMKQFSENIPTELLSAKIDGCGEIRTFINVAFFIVKPG 194  
G+ ++LW +I+P V P G+F+ Q+ ++IP ELLESAKIDG E + F + FP+ KP  
Sbjct: 129 GLINSLWGLIIPAVYTPGTGMFMAVQYMKDIPDELLESAKIDGANEWQIFWRIVFPLSKPL 188

20 Query: 195 FAALAIFTTINTWNDYFMQLVMLTSRNLITISLGVATMQAEM-ATNYGLIMAGAALAAVP 253  
AALAIF+F WND+ + L+++ RN T+ L +AT+Q E + I+A + L +P  
Sbjct: 189 VAALAIFSTFRWVDFVLPPLLVNRRNLTYLQLALATIQEYGGAEWNTILAFSTLTIIIP 248

Query: 254 IVTVFLVFQKSFTQGITMGAVK 275  
+ +FL+FQ+ F +GI G +K  
Sbjct: 249 TLIIFLLFQRLFMKGIMAGGLK 270

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 937> which encodes the amino acid sequence <SEQ ID 938>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have a cleavable N-term signal seq.

30 INTEGRAL Likelihood = -6.37 Transmembrane 245 - 261 ( 240 - 265)  
INTEGRAL Likelihood = -5.15 Transmembrane 140 - 156 ( 139 - 158)  
INTEGRAL Likelihood = -2.97 Transmembrane 111 - 127 ( 107 - 128)  
INTEGRAL Likelihood = -2.87 Transmembrane 76 - 92 ( 75 - 93)  
INTEGRAL Likelihood = -1.59 Transmembrane 188 - 204 ( 186 - 204)

35 ----- Final Results -----

bacterial membrane --- Certainty=0.3548(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the databases:

>GP:CAB59597 GB:AL132662 probable sugar transport inner membrane  
protein [Streptomyces coelicolor A3(2)]  
Identities = 88/262 (33%), Positives = 147/262 (55%)

45 Query: 15 VMLCVLTILFIFFPYWIMTGAFKAQADTIMIPPQWWPKAPTIENTFKALVVQNPAIKLWN 74  
++L L ++F P W++ + + A+ PP WP + ++ ++ +W N  
Sbjct: 38 LLLAPLALVFAVPLVWLVLSSVMSNAEINRFPALWPSGIDLGGRYVVLGNAMFPRWFVN 97

50 Query: 75 SVFISVATMFLVCGTSSLAGYALAKKRFYQRLIFSIFIAAMALPKQVVLVPLVRIVNFM 134  
S+ +S T+ SLAGYA A+ RF G R+L + +A MA+P Q+ ++P ++ +  
Sbjct: 98 SLIVSAVTVAANLVFGSLAGYAFARMRFAGSRVLMGLMLATMAVPFQLTMIPTFLVMKKL 157

55 Query: 135 GIHDTLAAVILPLVGWPFVFLMKQFSENIPTELLSAKIDGCGEIRTFINVAFFIVKPG 194  
G+ DTL A+I+P + PF VFL++QF ++P EL E+A IDGC +R + + P+ +P  
Sbjct: 158 GLIDTLGALIVPSLVTPFAVFLRQFFLSLPRELEEAAWIDGCSRLRVLWRIVLPLSRPA 217

60 Query: 195 FAALAIFTTINTWNDYFMQLVMLTSRENLTISLGVATMQAEMATNYGLIMAGAAMAAPVI 254  
A +A+ TF+ TWND L+ + T+ LG+ T Q + T + +MAG + +P+  
Sbjct: 218 LATVAVLTFLTWTWDLTWPLIAINHDTQYTLQLGLTTFQGHHTQWAAVMAGNVITVLPV 277

Query: 255 VTVFLVFQKSFTQGITMGAVKG 276  
+ FL QK+F Q IT +KG  
Sbjct: 278 LLAFLGAQKTFIQSITSSGLKG 299

65 An alignment of the GAS and GBS proteins is shown below:

-370-

Identities = 245/276 (88%), Positives = 262/276 (94%)

Query: 1 MKKKTFSAYNFLTALILCLLTVLFIFFPYWIMTGAFKSPDTIIPPQWWPKAPTLENFK 60  
 M KK +A + LT ++LC+LT+LFIFPFYWIMTGAFK+Q DTI+IPPQWWPKAPT+ENFK  
 5 Sbjct: 1 MTKKKLTASDILTTVMCLVLTILFIFFPYWIMTGAFKAQADTIMIPPQWWPKAPTLENFK 60

Query: 61 ALTVQNPALRWLWNSVFISIMTMFLVCCTSSMAGYVLAKKRFYQKILFSLFIAAMALPK 120  
 AL VQNPAL+WLWNSVFIS+ TMFLVC TSS+AGY LAKKRFYQ++LFS+FIAAMALPK  
 10 Sbjct: 61 ALVVQNPALKWLWNSVFISVATMFLVCGTSSLAGYALAKKRFYQRLLSIFIAAMALPK 120

Query: 121 QVVLVPLVRIINFMGIHDTLWAVILPLVGWPFVFLMKQFSENIPTELLESADKIDGCGEI 180  
 QVVLVPLVRI+NFMGIHDTL AVILPLVGWPFVFLMKQFSENIPTELLESADKIDGCGEI  
 15 Sbjct: 121 QVVLVPLVRIVNFMIHDTLAAVILPLVGWPFVFLMKQFSENIPTELLESADKIDGCGEI 180

Query: 181 RTFINVAFPIVKPGFAALAIFFINTWNDYFMQLVMTSRNLTISLGVATMQAEMATNY 240  
 RTF NVAFPIVKPGFAALAIFFINTWNDYFMQLVMTSR NLTISLGVATMQAEMATNY  
 20 Sbjct: 181 RTFFNVAFPIVKPGFAALAIFFINTWNDYFMQLVMTSRNLTISLGVATMQAEMATNY 240

Query: 241 GLIMAGAALAAVPIVTVFLVFQKSFTQGITMGA VKG 276  
 GLIMAGAA+AAVPIVTVFLVFQKSFTQGITMGA VKG  
 25 Sbjct: 241 GLIMAGAAMAAPVPIVTVFLVFQKSFTQGITMGA VKG 276

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 25 Example 292

A DNA sequence (GBSx0320) was identified in *S.agalactiae* <SEQ ID 939> which encodes the amino acid sequence <SEQ ID 940>. Analysis of this protein sequence reveals the following:

Possible site: 31  
 >>> Seems to have a cleavable N-term signal seq.

30	INTEGRAL	Likelihood = -10.83	Transmembrane	74 - 90 ( 64 - 96)
	INTEGRAL	Likelihood = -6.37	Transmembrane	108 - 124 ( 107 - 126)
	INTEGRAL	Likelihood = -5.84	Transmembrane	270 - 286 ( 265 - 290)
	INTEGRAL	Likelihood = -5.20	Transmembrane	161 - 177 ( 156 - 182)
35	INTEGRAL	Likelihood = -0.16	Transmembrane	219 - 235 ( 219 - 235)

----- Final Results -----  
 bacterial membrane --- Certainty=0.5331(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 40 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB05584 GB:AP001513 sugar transport system (permease) (binding protein dependent transporter) [Bacillus halodurans]  
 Identities = 106/289 (36%), Positives = 168/289 (57%), Gaps = 6/289 (2%)

45 Query: 9 RETMIAYAFLAPILLFFLIFVFAPVMVGFTSFFNYSM-TQFTFIGLANYNRMF-HDSIF 66  
 +E Y F+AP ++ F IF PM+ SF ++ + + + G NY R+F D +F  
 Sbjct: 25 KEYFWGYLFIAPPIIGFAIFALGPMLYSIYVSFTDFDLYNEPVWTGADNYYRLFVTDLDL 84

50 Query: 67 MKSLINTVILVIGSVPVVVFSLFVAANTYKENVFSRSFYRCVFFLPVVTGSAVTVVWK 126  
 K++ NT +G +P+ + SL +A +K V + +R FFLP V+ VA+T++W+  
 Sbjct: 85 RKTVFNTFFYAALG-IPIGMAVSLGIAVALNQK-VKGIALFRTAFFLPAVSSVVAITLLWR 142

55 Query: 127 WIYDPMGILNYILKSGHVIEQNISWLGDKHWALLAIIIIILLTTSVGQPIILYIAAMGNI 186  
 WI++ G+LN +L +V WL D+ WA+ A+II + +G +ILY+AA+ +  
 Sbjct: 143 WIFNADFGLNIMLN--YVGIHGPGLSDEKWAMPAMIIQGVWGLGINMILYLAALQGV 200

60 Query: 187 DNSLCEAARVDGANEMQVFWQIKWPSLLPTTLYIAVITTINSFQCFALIQLLTSGGPNYS 246  
 + +L EAA +DG N Q F I PS+ PTT +I + +TI + Q F ++T GGPNNYS  
 Sbjct: 201 NPALYEAADIDGGNAWQKFIHITVPSISPTTFFILITSTIGALQDFQRFMIMTEGGPNYS 260

Query: 247 TSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIALISFAQFKILGNDVEY 295

-371-

T+T++YYL+ AF+ E GYA+ M L ++I +I+ FK+ V Y  
 Sbjct: 261 TTTVVYYLFLNAFRYMEMGYASAMAWVLGIIILIIITIINFKLAKKWVHY 309

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 941> which encodes the amino acid  
 5 sequence <SEQ ID 942>. Analysis of this protein sequence reveals the following:

Possible site: 13  
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -12.74	Transmembrane	55 - 71 ( 44 - 78)
INTEGRAL	Likelihood = -10.83	Transmembrane	109 - 125 ( 98 - 130)
INTEGRAL	Likelihood = -6.21	Transmembrane	304 - 320 ( 299 - 324)
INTEGRAL	Likelihood = -6.00	Transmembrane	142 - 158 ( 141 - 160)
INTEGRAL	Likelihood = -5.04	Transmembrane	196 - 212 ( 190 - 216)
INTEGRAL	Likelihood = -0.16	Transmembrane	253 - 269 ( 253 - 269)

----- Final Results -----  
 bacterial membrane --- Certainty=0.6095(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20 The protein has homology with the following sequences in the databases:

>GP:BAB05584 GB:AP001513 sugar transport system (permease) (binding  
 protein dependent transporter) [Bacillus halodurans]  
 Identities = 113/310 (36%), Positives = 176/310 (56%), Gaps = 9/310 (2%)

Query: 25 KVEQKKEVFQVNVNKLKMR---ETLISYAFLAPVLVFFVIFVLIPMIMGFVTSFFNYSM- 80  
 +VE +E K K R E Y F+AP ++ F IF L PM+ SF ++ +  
 Sbjct: 4 EVETPRETKTKARKQKRRLNKEYFWGYLFIAPPTIGFAIFALGPMLYSTIYVSFTDFDLY 63

Query: 81 TEFTFVGFANYARMF-QDPIFMKSLINTLIIVIGSVPVVFFSLFVAAKTYDKNVVARSF 139  
 E + G NY R+F D +F K++ NT +G +P+ + SL +A K V +  
 Sbjct: 64 NEPVWTGADNYRFLVTDLFRKTVFNTFYAALG-IPIGMAVSLGIAVALNQK-VKGIAL 121

Query: 140 YRAVFFLPVVTGSAVTVVWKWIYDPMGILNYVLKYAHVIEQNISWLGDKHWALLAIIV 199  
 +R FFLP V+ VA+T++W+WI++ G+LN +L Y + WL D+ WA+ A+I+  
 Sbjct: 122 FRTAFLPAVSSVAITLLWRWIFNADFGLLNIMLNIVGI--HGPGWLSDEKWAMPAMII 179

Query: 200 ILLTTSVGQPIILYIAAMGNIDNSLVEAARVDGATEFQVFNWIKWPSLLPTTLYIAVITT 259  
 + +G +ILY+AA+ ++ +L EAA +DG +Q F +I PS+ PTT +I + +T  
 Sbjct: 180 QGVWGLGINMILYLAALQGVNPALYEAADIDGGNAWQKFIHITVPSISPTTFFILITST 239

Query: 260 INSFQCFALIQLLTSGGPNYSTSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIAIISFAQ 319  
 I + Q F ++T GGPNYST+T++YYL+ AF+ E GYA+ M L ++I II+  
 Sbjct: 240 IGALQDFQRFMIMTEGGPNYSTTTVVYYLFLNAFRYMEMGYASAMAWVLGIIILIIITIIN 299

Query: 320 FKILGNDVEY 329  
 FK+ V Y  
 Sbjct: 300 FKLAKKWVHY 309

An alignment of the GAS and GBS proteins is shown below:

Identities = 263/295 (89%), Positives = 278/295 (94%)

Query: 1 MRTNKLKMRMTMIAYAFLAPILLFFLIFVFAPVMVGMGFVTSFFNYSMQTFTFIGLANYNRM 60  
 + NKLKMRRET+I+YAFLAP+L+FF+IFV PM+MGFVTSFFNYSM+FTF+G ANY RM  
 Sbjct: 35 VNVNKLKMRRET LISYAFLAPVLVFFVIFVLIPMIMGFVTSFFNYSMTEFTFVGFANYARM 94

Query: 61 FHDSIFMKSLINTVIIVIGSVPVVFFSLFVAANTYEKNVFSRSFYRCVFFLPVVTGSA 120  
 F D IFMKSLINT+IIVIGSVPVVFFSLFVAA TY+KNV +RSFYR VFFLPVVTGSA  
 Sbjct: 95 FQDPIFMKSLINTLIIVIGSVPVVFFSLFVAAKTYDKNVVARSFYRAVFFLPVVTGSA 154

Query: 121 VTVVWKWIYDPMGILNYILKSGHVIEQNISWLGDKHWALLAI IILLTTSVGQPIILYI 180  
 VTVVWKWIYDPMGILNY+LK HVIEQNISWLGDKHWALLAI IILLTTSVGQPIILYI  
 Sbjct: 155 VTVVWKWIYDPMGILNYVLKYAHVIEQNISWLGDKHWALLAI IILLTTSVGQPIILYI 214

Query: 181 AAMGNIDNSLCEAARVDGANEMQVFWQIKWPSLLPTTLYIAVITTINSFQCFALIQLLTS 240

-372-

AAMGNIDNSL EAARVDGA E QVFW IKWPSLLPTTLYIAVITTINSFQCFALIQLLTS  
 Sbjct: 215 AAMGNIDNSLVEAARVDGATEFQVFNKIKWPSLLPTTLYIAVITTINSFQCFALIQLLTS 274

Query: 241 GGPNYSTSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIALISFAQFKILGNDVEY 295  
 GGPNYSTSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIA+ISFAQFKILGNDVEY  
 Sbjct: 275 GGPNYSTSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIAIISFAQFKILGNDVEY 329

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 10 Example 293

A DNA sequence (GBSx0321) was identified in *S.agalactiae* <SEQ ID 943> which encodes the amino acid sequence <SEQ ID 944>. Analysis of this protein sequence reveals the following:

Possible site: 31  
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----  
                   bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12516 GB:Z99107 similar to sugar-binding protein [Bacillus subtilis]  
 Identities = 54/187 (28%), Positives = 90/187 (47%), Gaps = 14/187 (7%)

Query: 19 MFACVDSSQSVMAAEKD-KVEITWWAFPTFTQEKAKDGVGTYEKKVIKAFKKNPNKVKV 77  
 MF+ + + ++D + I WW + D Y KVI+ +EKKNP++ ++  
 Sbjct: 1 MFSGCSAGEEASGKKEDVTLRIAWWG-----GQPRHD----YTTKVIELYEKKNPVHIE 51

Query: 78 LETIDFTSGPEKITTAIEAGTAPDVLFDAPGRIIQYGKNGKLADLNDLFTDQFIKDVN-- 135  
 E ++ +K+ AG PDV+ + QYGK +L DL D I DV+  
 Sbjct: 52 AEFANWDDYWKKLAPMSAAGQLPDVIQMDTAYLAQYGKKNQLEDLTPYTKDGTI-DVSSI 110

Query: 136 NKNIIQASKSGDKAYMPISSAPFYMAFNKKMLKDAGVLKLVKEGWTTSDFEKVLKALKN 195  
 ++N++ K +K Y + + + N+ +LK AGV + +E WT D+EK+ L+  
 Sbjct: 111 DENMLSGGKIDNKLYGFTLGVNVLSVIANEDLLKKAGV-SINQENWTWEDYEKLYADLQE 169

Query: 196 KGYTPGS 202  
 K GS  
 Sbjct: 170 KAGVYGS 176

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 945> which encodes the amino acid sequence <SEQ ID 946>. Analysis of this protein sequence reveals the following:

Possible site: 20  
 >>> May be a lipoprotein

----- Final Results -----  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

!GB:Z99107 similar to sugar-binding protein [Bacillu... 82 2e-14

>GP:CAB12516 GB:Z99107 similar to sugar-binding protein [Bacillus subtilis]  
 Identities = 105/446 (23%), Positives = 176/446 (38%), Gaps = 71/446 (15%)

Query: 24 GKSQKEAGASKSDTAKTEITWWAFPVFTQEKABDGVGTYEKKLIAAFEKANPEIKVKLET 83  
 G S E + K + I WW + D Y K+I +EK NP + ++ E

-373-

Sbjct: 4 GCSAGEEASGKKEDVTLRIAWWG-----GQPRHD-----YTTKVIELYEKKKNPHVHIEAEF 54

Query: 84 IDFTSGPEKITTAIEAGTAPDVLFDAPGRRIQYGKNGKLADLNDLFTEFTKDVN--NDK 141  
 ++ +K+ AG PDV+ + QYGK +L DL +T++ T DV+ ++

5 Sbjct: 55 ANWDDYWKKLAPMSAAGQLPDVIQMDTAYLAQYGKKNQLEDLTP-YTKDGTIDVSSIDEN 113

Query: 142 LIQASKAGDTAYMYPISAPFYMALNKKMLKDAGVLDLVKEGWTTDDFEKVLKALKDK-- 199  
 ++ K + Y + + + N+ +LK AGV + +E WT +D+EK+ L++K

10 Sbjct: 114 MLSSGGKIDNKLYGFTLG VNVLSVIANEDLLKKAGV-SINQENWTWEDYEKLAYDLQEKAG 172

Query: 200 -----GYNPGSFFFANGQGGDQGPRAFFANLYSSHITDDKV-----TKYTT 239  
 G +P F +G R + + DD++ T T

15 Sbjct: 173 VYGSNGMHPPDIFFPYYLRTKGERFYKEDGTGLAYQDDQLFVDYFERQLRLVKAKTSPTP 232

Query: 240 DDANSIKAMTKISNWKDGLMMNGSQYDGSADIQNFANGQTSFTILWAPAQPGIQAKLLE 299  
 D++ IK M +D ++ G SA N++N F A+L +

20 Sbjct: 233 DESAQIKGM-----EDDFIVKGGK----SAITWYNSNQYLGK-----ARLTD 269

Query: 300 ASKVDYLEIPFSDDGKPELEYLVNGFAVFNNDKDEQKVAASKTFIQFIADDDKEWGPKNVV 359  
 + YL P + L + E K A+K FI F +++E + +

25 Sbjct: 270 SPLSLYLP--PEQMGEKALTLPKSMFLSIPKSSEHKKEAAK-FINFFVNNEE-ANQLIK 324

Query: 360 RTGAPFVRTSYGDLYKDKRMEK---IAEWTKFYSPYNTID-----GFAEMRTLWFFPMVQ 411  
 PV D K K E+ I E+ + S + D G AE+ L

30 Sbjct: 325 GERGVPSVDKVAIDAIKPKLNEEETNIVEYVETASKNISKADPPEPVGSAEVIKLLKDTSD 384

Query: 412 AVSNGDEKPEDALKAFTEKANKTIKK 437  
 + PE A K F +KAN+ +++

Sbjct: 385 QILYQKVSPEKAAKTRKKANEILER 410

An alignment of the GAS and GBS proteins is shown below:

Identities = 352/438 (80%), Positives = 384/438 (87%), Gaps = 4/438 (0%)

35 Query: 1 MSIKKSVIGFCLGAAALSMFACVDSSQSVMAAEKD---KVEITWWAFPTFTQEAKADGVG 57  
 M++KK LGA+ L + AC SQ A K K EITWWAFP FTQEKA+DGVG

Sbjct: 1 MNMKKLASLAMLGASVGLAACGGKSQKEAGASKSDTAKTEITWWAFPVFTQEKAEDGVG 60

40 Query: 58 TYEKKVIKAFEEKNPNIKVKLETIDFTSGPEKITTAIEAGTAPDVLFDAPGRRIQYGKNG 117  
 TYEKK+I AFEK NP IKVKLETIDFTSGPEKITTAIEAGTAPDVLFDAPGRRIQYGKNG

45 Sbjct: 61 TYEKKLIAAFEKANPEIKVKLETIDFTSGPEKITTAIEAGTAPDVLFDAPGRRIQYGKNG 120

Query: 118 KLADLNDLFTDQFIKDVNNKNIIQASKSGDKAYMYPISAPFYMAFNKMLKDAGVLKLV 177  
 KLADLNDLFT++F KDVNN +IQASK+GD AYMPISSAPFYMA NKKMLKDAGVL LV

50 Sbjct: 121 KLADLNDLFTEFTKDVNDKLIQASKAGDTAYMYPISAPFYMALNKKMLKDAGVLDLV 180

Query: 178 KEGWTTSDFEKVLKALKKNKGYTPGSFFANGQGGDQGPRAFFANLYSAPITDKEVTKYTTD 237  
 KEGWTT DFEKVLKALK+KGY PGSFFANGQGGDQGPRAFFANLYS+ ITD +VTKYTTD

55 Sbjct: 181 KEGWTTDDFEKVLKALKDKGYNPGSFFANGQGGDQGPRAFFANLYSSHITDDKVTKYTTD 240

Query: 238 TKNSVKSMKKIVEWIKKGYLMNGSQYDGSADIQNFANGQTAFTILWAPAQPKTQAKLLES 297  
 NS+K+M KI WIK G +MNGSQYDGSADIQNFANGQT+FTILWAPAQP QAKLLE+

60 Sbjct: 241 DANSIKAMTKISNWKDGLMMNGSQYDGSADIQNFANGQTSFTILWAPAQPGIQAKLLEA 300

Query: 298 SKVDYLEVPFPPSEDGKPDLEYLVNGFAVFNNDKENVKASKKFITFIADDDKKGPKDVIR 357  
 SKVDYLE+PFP+DGKP+LEYLVNGFAVFNNDK KV ASK FI FIADDDK+WGPK+V+R

65 Sbjct: 301 SKVDYLEIPFSDDGKPELEYLVNGFAVFNNDKDEQKVAASKTFIQFIADDDKEWGPKNVVR 360

Query: 358 TGAFPPVRTSFGDLYKDKRMMKISKWTQYSPYNTIDGFSEMRTLWFFPMVQSVSNGDEK 417  
 TGAFPPVRTS+GDLYK DKRM KI++WT++YSPYNTIDGF+EMRTLWFFPMVQ+VSNGDEK

Sbjct: 361 TGAFPPVRTSYGDLYK-DKRMEKIAEWTKFYSPYNTIDGFAEMRTLWFFPMVQAVSNGDEK 419

Query: 418 PADALKDFTQKANDTIKK 435  
 P DALK FT+KAN TIKK

Sbjct: 420 PEDALKAFTEKANKTIKK 437

```

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-----SPEKAAKTERKKANEILERNN



SEQ ID 944 (GBS16) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 3 (lane 9; MW 49kDa).

The GBS16-His fusion product was purified (Figure 92A; see also Figure 189, lane 9) and used to immunise mice (lane 1 + 2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 92B), FACS (Figure 92C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 294

10 A DNA sequence (GBSx0322) was identified in *S.agalactiae* <SEQ ID 947> which encodes the amino acid sequence <SEQ ID 948>. Analysis of this protein sequence reveals the following:

Possible site: 49  
>>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20 A related GBS nucleic acid sequence <SEQ ID 9459> which encodes amino acid sequence <SEQ ID 9460> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC66999 GB:AE001166 conserved hypothetical protein [Borrelia burgdorferi]  
25 Identities = 107/225 (47%), Positives = 147/225 (64%), Gaps = 6/225 (2%)  
  
Query: 12 QIKNGIIVSCQALPGEPLYTESGGVMPDLLALAAQEAGAVGIRANSVRDIKEIQEVTNLPI 71  
+IK G+IVSCQAL EPL+ S +M +ALAA+ GA+GIRAN V DI +I+ +LPI  
Sbjct: 6 KIKRGLIVSCQALENEPLH--SSFIMSKMALAAKIGGAIGIRANGVNDISQIKLEVLDPI 63  
  
30 Query: 72 IGIKREYPPQEPFITATMTEVDQLASLDIAVIALDCTLRERHDGLSVVEFIQIKRKYP 131  
IGIIK+ Y + FIT TM E+D+L + + +IALD T R R DG+ + +F + IK+KYP  
Sbjct: 64 IGIKKNYNNCDVFITPTMKEIDELCNEGVDIIALDATFRNRPDGVLLDDFFENIKKKYP 123  
  
35 Query: 132 EQLLMADISTFEEGKNAFEAGVDFVGTTLGTYDYSR--QEEGPDIELLNKLCQAGI--D 187  
+Q LMADIS+ +E NA + G DF+GTTL GYT + D L L + +  
Sbjct: 124 KQCLMADISSLDEAINADKLGDFDFIGTTLGYTKNTNGLNIADNDFNFLRTLNSNLKST 183  
  
40 Query: 188 VIAEGKIHTPKQANEINHIGVAGIVVGGAITRPKEIAERFISGLS 232  
+I EGKI TP +A + +GV +VVGGAITRP EI ++F+ ++  
Sbjct: 184 LIVEGKIDTPLKAQKCFEMGVDLVVVGGAITRPAEITKKFVEKIN 228

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 949> which encodes the amino acid sequence <SEQ ID 950>. Analysis of this protein sequence reveals the following:

45 Possible site: 44  
>>> Seems to have no N-terminal signal sequence  
INTEGRAL Likelihood = -1.49 Transmembrane 175 - 191 ( 175 - 192)  
  
50 ----- Final Results -----  
bacterial membrane --- Certainty=0.1595(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-376-

The protein has homology with the following sequences in the databases:

```
>GP:AAD28762 GB:AF130859 putative N-acetylmannosamine-6-P epimerase
[Clostridium perfringens]
Identities = 113/225 (50%), Positives = 148/225 (65%), Gaps = 5/225 (2%)

5   Query: 10  LMEQLKGGIIVSCQALPGEPLYSETGGIMPLMAKAAQEAGAVGIRANSVRDIKEIQAITD 69
      +++ +KG +IVSCQAL EPL+S IM MA AA++ GA IRA + DI EI+ +T
      Sbjct: 1  MLDVVKGNLIVSCQALSDEPLHSSF--IMGRMAIAAKQGGAIRAQGIDDINEIKEVTK 58

10  Query: 70  LPIIGIHKDYPPQEPFITATMTEVDQLAALNIAVIAMDCTKRDRHDGLDIASFIRQVKE 129
      LPIIGIHK++Y E +IT TM EVD+L + +I +D TKR R +G +I + +
      Sbjct: 59  LPIIGIHKRNYDDSEIYITPTMKEVDELLKTDCEMIGLDATKRKRPNGENIKDLVDAlHA 118

15  Query: 130  KYPNQLLMADISTFDEGLVAHQAGIDFVGTTLSGYTPYSRQEAGPDVALIEALCK-AGIA 188
      K +L MADIST +EG+ A + G D V TTLSGYTPYS+Q D L+E L K I
      Sbjct: 119 K--GRLAMADISTLEEGIEAEKLGFDGVSTTTLSGYTPYSKQNSVDFELLEELVKTVPKIP 176

      Query: 189 VIAEGKIHSPEEAKKINDLGAVIGVGGAITRPKEIAERFIEALK 233
      VI EG+I++PEE KK DLG VVGGAITRP++I +RF + LK
20  Sbjct: 177 VICEGRINTPEELKKALDLGAYSAVVGGAITRPQQITKRFTDILK 221
```

An alignment of the GAS and GBS proteins is shown below:

Identities = 172/227 (75%), Positives = 202/227 (88%)

```
25  Query: 5  SKEAFKKQIKNGIIVSCQALPGEPLYTESGGVMPLLAQAQEAGAVGIRANSVRDIKEIQ 64
      +KE +Q+K GIIIVSCQALPGEPLY+E+GG+MPL+A AAQEAGAVGIRANSVRDIKEIQ
      Sbjct: 6  TKEKLMEQLKGGIIVSCQALPGEPLYSETGGIMPLMAKAAQEAGAVGIRANSVRDIKEIQ 65

30  Query: 65  EVTNLPIIGIHKREYPPQEPFITATMTEVDQLASLDIAVIALDCTLRERHDGLSVVEFIQ 124
      +T+LPIIGIHK++YPPQEPFITATMTEVDQLA+L+IAVIA+DCT R+RHDGL + FI+
      Sbjct: 66  AITDLPIIGIHKDYPPQEPFITATMTEVDQLAALNIAVIAMDCTKRDRHDGLDIASFIR 125

      Query: 125 KIKRKYPEQLLMADISTFEEGKNAFEAGVDFVGTTLSTGYTDYSRQEEGPDIELLNKLCQA 184
      ++K KYP QLLMADISTF+EG A +AG+DFVGTTLSTGYT YSRQE GPD+ L+ LC+A
35  Sbjct: 126 QVKEKYPNQLLMADISTFDEGLVAHQAGIDFVGTTLSGYTPYSRQEAGPDVALIEALCKA 185

      Query: 185 GIDVIAEGKIHTPKQANEINHIGVAGIVVGGAITRPKEIAERFISGL 231
      GI VIAEGKIHP++A +IN +GVAGIVVGGAITRPKEIAERFI L
40  Sbjct: 186 GIAVIAEGKIHSPEEAKKINDLGAVIGVGGAITRPKEIAERFIEAL 232
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 295

A DNA sequence (GBSx0323) was identified in *S.agalactiae* <SEQ ID 951> which encodes the amino acid  
 45 sequence <SEQ ID 952>. This protein is predicted to be group B streptococcal surface immunogenic protein. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have a cleavable N-term signal seq.
```

```
50  ----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

55 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 953> which encodes the amino acid sequence <SEQ ID 954>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
```

-377-

&gt;&gt;&gt; Seems to have a cleavable N-term signal seq.

----- Final Results -----

5                   bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10           Identities = 182/437 (41%), Positives = 240/437 (54%), Gaps = 53/437 (12%)

Query: 1   MKMKNKVLTLTSTMAASLLSVASVQAQETDTTWTARTVSEVKADLVKQDNKSSYTVKYGDT 60  
           M + KK L +++A SL+ +A+ QAQE   WT R+V+E+K++LV   DN +YTVKYGDT  
 Sbjct: 1   MIITKKSFLVTSVALSLVPLATAQAQE----WTPRSVTEIKSELVLVDNVFTYTVKYGDT 56

15           Query: 61   LSVISEAMSIDMNVLAKEINNIADINLIYPETTLTVTYDQKSHATSMKIETPATNAAGQT 120  
                   LS I+EAM ID++VL   IN+IA+I+LI+P+T LT   Y+Q   AT++ ++ PA++ A +  
 Sbjct: 57   LSTIAEAMGIDVHVLGDINHIANIDLIFPDILTANYNQHGQ-ATNLTVQAPASSPASVS 115

20           Query: 121 TATVDLKTNQVSADQKVSLNTISEGMTP-EAATTIVSPMKTYSSAPALKSKEVLAQEQA 179  
                   Q S Q           ++ TP + TT + K SS A S E + +  
 Sbjct: 116 HVPSSSEPLQASATSQPTV--PMAPPATPSDVPTTFFASAKPDSSVTA--SSELTSTND 171

25           Query: 180 VSQAAANEQVSPAPVKSITSEVPAAKEEVKPTQTSVSQSTTVSPASVAAETPAPVAKVAP 239  
                   VS ++E       V       P A E       T V   T +S A +A P P   +  
 Sbjct: 172 VSTELSSSESQKQPEVPQEA VPTPKAAE-----TTEVEPKTDISEAPTSANRPVPNESASE 226

30           Query: 240 VRTVAAPRVASVKVVTPKVETGASPEHVSAPAVP---VTTTSPATDSKLQATEVKSVPVA 296  
                   + AAP                   + A E   SAPA       TTS AT + L  
 Sbjct: 227 EVSSAAP-----AQAPAEKEETSAPAAQKAVADTTSVATSNGL----- 264

35           Query: 297 QKAPTATPVAQPASTTNAVAHPENAGLQPHVAAAYKEKVASTYGVNEFSTYRAGDPGDHG 356  
                   AP                   A +P NAGLQP   AA+KE+VAS +G+   FS YR GDPGDHG  
 Sbjct: 265 SYAPNH-----AYNPMNAGLQPTA AFKEEVASAFGITSFSGYRPGDPGDHG 311

40           Query: 357 KGLAVDFIVGTNQA LGNKVAQYSTQNMAANNISYVIWQKQFYSNTNSIYGPANTWNAMPD 416  
                   KGLA+DF+V N ALG++VAQY+ +MA   ISYVIW+Q+FY+   SIYGPA TWN MPD  
 Sbjct: 312 KGLAIDFMVPENSALGDQVAQY AIDHMAERGISYVIWKQRFYAPFASIYGPAYTWNMPD 371

40           Query: 417 RGGVTANHYDHHVHVSFN 433  
                   RG +T NHYDHHVHVSFN  
 Sbjct: 372 RGSITENHYDHHVHVSFN 388

A related GBS gene <SEQ ID 8539> and protein <SEQ ID 8540> were also identified. Analysis of this protein sequence reveals the following:

45   Lipop: Possible site: -1   Crend: 3  
      SRCFLG: 0  
      McG: Length of UR: 20  
           Peak Value of UR: 1.96  
           Net Charge of CR: 2

50   McG: Discrim Score: 2.95  
      GvH: Signal Score (-7.5): 3.84  
           Possible site: 23

>>> Seems to have a cleavable N-term signal seq.  
 Amino Acid Composition: calculated from 24

55   ALOM program   count: 0 value: 4.29 threshold: 0.0  
      PERIPHERAL Likelihood = 4.29       58  
      modified ALOM score: -1.36

\*\*\* Reasoning Step: 3

60   Rule gpol

----- Final Results -----

65                   bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-378-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

SEQ ID 8540 (GBS322) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 9; MW 52kDa). The GBS322-His fusion product was purified (Figure 214, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 267), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 296**

A DNA sequence (GBSx0324) was identified in *S.agalactiae* <SEQ ID 955> which encodes the amino acid sequence <SEQ ID 956>. Analysis of this protein sequence reveals the following:

Possible site: 23

&gt;&gt;&gt; Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -1.86 Transmembrane 5 - 21 ( 4 - 21)

----- Final Results -----

bacterial membrane --- Certainty=0.1744 (Affirmative) &lt; succ&gt;

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC46072 GB:U50357 zoocin A endopeptidase [Streptococcus  
zooeptidemicus]

Identities = 163/274 (59%), Positives = 196/274 (71%), Gaps = 11/274 (4%)

Query: 25 VLADTYVRPIDNGRITTFGNGYPGHCGVDYAVPTGTIIRAVADGTVKFAGAGANFSWMTD 84  
V A TY RP+D G ITTGFGNGYPGH GVDYAVP GT +RAVA+GTVKFAG GAN WM

Sbjct: 21 VSAATYTRPLDTGNTTTFGNGYPGHVGVDYAVPVGTPVRAVANGTVKFAGNGANHPWMLW 80

Query: 85 LAGNCVMIQHADGMHSGYAHMSRVVARTGEKVKQGDIIIGYVGATGMATGPHLHFEFLPAN 144  
+AGNCV+IQHADGMH+GYAH+S++ T VKQG IIGY GATG TGPHLHFE LPAN

Sbjct: 81 MAGNCVLIQHADGMHTGYAHLISKISVSTDSTVKQGGIIGYTGATGQVTGPHLHFEMLPAN 140

Query: 145 PNFQNGFHFGRINPTSLIANVATFSGKTQASAPSIKPLQSAAPVQNQSSKLKVYRVDELQKV 204  
PN+QNGF GRI+PT IAN F+G T + P N LK+Y+VD+LQK+

Sbjct: 141 PNWQNGFSGRIDPTGYIANAPVFNGTTPTPE-----PTTPTTN----LKIKVDDLQKI 189

Query: 205 NGVWLVKNNLTPTGFDWNDNGIPASEIDEVDANGNLTDQVLQKGGYFIFNPKTLKTVE 264  
NG+W V+NN L PT F W DNGI A ++ EV +NG T+DQVLQKGGYF+ NP +K+V

Sbjct: 190 NGIWQVRNNILVPTDFTWVDNGIAADDVIEVTSNGTRTSDQVLQKGGYFVINPNNVKSVM 249

Query: 265 KPIQGTAGLTWAKTRFANGSSVWLRVDNSQELLY 298  
P++G+ GL+WA+ F G +VWL + LLY

Sbjct: 250 TPMKSGGLSWAQVNFTTGGNVWLNTTSKDNLLY 283

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8541> and protein <SEQ ID 8542> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 6

McG: Discrim Score: 6.63

GvH: Signal Score (-7.5): -2.97

Possible site: 23

&gt;&gt;&gt; Seems to have an uncleavable N-term signal seq

ALOM program count: 1 value: -1.86 threshold: 0.0

INTEGRAL Likelihood = -1.86 Transmembrane 5 - 21 ( 4 - 21)

PERIPHERAL Likelihood = 5.57 50



-380-

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2815(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04352 GB:AP001509 phosphoribosylaminoimidazolecarboxamide

formyltransferase/IMP cyclohydrolase [Bacillus halodurans]

Identities = 310/515 (60%), Positives = 390/515 (75%), Gaps = 4/515 (0%)

10

Query: 1 MTKRALISVSDKSGIIDFAKELKNLGWDIISTGGTKVALDDAGVETIAIDVTGFPEMMD 60  
 M +RAL+SVS+K GI+ FAK L ++I+STGGTK AL +AG+ I DVTGFPE++D  
 Sbjct: 1 MKRRALVSVSNKEGIVPFAKALVEHEVEIVSTGGTKRALQEAGIPVTGISDVTGFPEILD 60

15

Query: 61 GRVKTLHPNIHGGLLARRDADSHLQAADNNIELIDLVVVNLYPFKETILRPDVTYDLAV 120  
 GRVKTLHPNIHGGLLA R+ D HL +++I ID VVVNLYPF++TI +P+ T+ A+  
 Sbjct: 61 GRVKTLHPNIHGGLLAMRERDEHLAQLNEHHIRPIDFVVVNLYPFQQTIAKPEATFADAI 120

20

Query: 121 ENIDIGGPSMLRSAAKNHASVTVVVDSADYATVLGELADASQTTFKTRQRLAAKAFRHTA 180  
 ENIDIGGPSMLR+AAKNH VVVVD DY TVL ELAD +T++RLAAK FRHTA  
 Sbjct: 121 ENIDIGGPSMLRAAAKNHQHVTVVVDVVDYETVLKELADQGNVATETKRRLLAAKVRHTA 180

25

Query: 181 AYDALIAEYFTAQVGEAKPEKLITITYDLKQAMRYGENPQQDADFQKALPTDYSIASAKQ 240  
 AYDA+IAEY T VGE PE LT+T++ KQ +RYGENP Q A FYQK L SIA AKQ  
 Sbjct: 181 AYDAMIAEYLTDAVGEESPESLTTFEKKQDLRYGENPHQKATFYQKPLGAKASIAHAKQ 240

30

Query: 241 LNKELSFNNIRDADAAIRIIRDFKDSPTVVALKHMNPGGIGQADDIETAWDYAYEADPV 300  
 L+GKELS+NNI DADAA+ I+++FK+ P VA+KHMNPGG+G + I+ A+D AYEADPV  
 Sbjct: 241 LHGKELSYNNINDADAALSIVKEFKE-PAAVAVKHMNPGVGTGETTKEAFDKAYEADPV 299

35

Query: 301 SIFGGIVVLNREVDATAEKMHPIFLEIIIIAPSYSEEALAILTNKKKNLRILELPFDAQA 360  
 SIFGGI+ LNREVD TA+ + IFLEIIIIAPS+SEEAL +LT+ KKNLR+L LP + +  
 Sbjct: 300 SIFGGIIALNREVDVETAKTLKEIFLEIIIIAPSFSEEALDVLTSS-KKNLRLLTLPLNEE- 357

40

Query: 361 ASEVEAEYTGTVGGLLVQNDQVVAENPSDWQVVTDRQPTQEATALEFANKAIKYVKSNG 420  
 ++ E T + GG LVQ +D ++ ++ T R+PTE E AL+ AW+ +K+VKS  
 Sbjct: 358 -NQAERITSIHGGALVQREDTYGFEEAEIKIPTKREPTAEWEALKLAWRVVVKHKSNA 416

Query: 421 IIIITNDHMTLGLGAGQTNRVGSVKIAIEQAKDHLDGAVLASDAFFPFADNIEEIAAGIK 480  
 I++ + MT+G+GAGQ NRVG+ KIAIEQA + G+V+ SDAFFP D +E A AGI  
 Sbjct: 417 IVLADGQMTVGVGAGQMNVRGAAKIAIEQAGEKAAGSVMGSDAFFPMGDTVELAAKAGIT 476

45

Query: 481 AIIQPGGSVRDQESIDAANKHGLTMIFTGVRHFRH 515  
 AIIQPGGS+RD+ESI+ A+KHG+ M+FTGVRHF+H  
 Sbjct: 477 AIIQPGGSIRDEESIENADKHGIAMVFTGVRHFKH 511

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 959> which encodes the amino acid sequence <SEQ ID 960>. Analysis of this protein sequence reveals the following:

Possible site: 48

50

>>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2932(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55

An alignment of the GAS and GBS proteins is shown below:

Identities = 500/515 (97%), Positives = 507/515 (98%)

60

Query: 1 MTKRALISVSDKSGIIDFAKELKNLGWDIISTGGTKVALDDAGVETIAIDVTGFPEMMD 60  
 MTKRALISVSDKSGI+DFAKELKNLGWDIISTGGTKV LDDAGVETIAIDVT FPEMMD  
 Sbjct: 1 MTKRALISVSDKSGIVDFAKELKNLGWDIISTGGTKVTLDDAGVETIAIDVTTRFPEMMD 60

Query: 61 GRVKTLHPNIHGGLLARRDADSHLQAADNNIELIDLVVVNLYPFKETILRPDVTYDLAV 120

-381-

GRVKTLHPNIHGGLLARRDADSHLQAAKDNNIELIDLIVVNLYPFKETILRPD+TYDLAV  
 Sbjct: 61 GRVKTLHPNIHGGLLARRDADSHLQAAKDNNIELIDLIVVNLYPFKETILRPDITYDLAV 120  
 Query: 121 ENIDIGGPSMLRSAAKNHASVTVVVDSADYATVLGELADASQTTFFKTRQRLAAKAFRHTA 180  
 ENIDIGGPSMLRSAAKNHASVTVVVD ADYATVLGELADA QTTFF+TRQRLAAK FRHTA  
 Sbjct: 121 ENIDIGGPSMLRSAAKNHASVTVVVDPADYATVLGELADAGQTTFFETRQRLAAKVFRTA 180  
 Query: 181 AYDALIAEYFTAQVGEAKPEKLTITYDLKQAMRYGENPQQDADFYQKALPTDYSIASAKQ 240  
 AYDALIAEYFT QVGEAKPEKLTITYDLKQAMRYGENPQQDADFYQKALPTDYSIASAKQ  
 Sbjct: 181 AYDALIAEYFTTQVGEAKPEKLTITYDLKQAMRYGENPQQDADFYQKALPTDYSIASAKQ 240  
 Query: 241 LNGKELSFNNIRDADAAIRIIRDFKDSPTTVVALKHMNPGGIGQADDIETAWDYAYEADPV 300  
 LNGKELSFNNIRDADAAIRIIRDFKD PTVVALKHMNPGGIGQADDIETAWDY Y+ADPV  
 Sbjct: 241 LNGKELSFNNIRDADAAIRIIRDFKDRPTTVVALKHMNPGGIGQADDIETAWDYTYKADPV 300  
 Query: 301 SIFGGIVVLNREVDAATAEKMHPIFLEIIIIAPSYSEEALAILTNKKKNLRILELPFDAQA 360  
 SIFGGI+VLNREVDAATA+KMHPIFLEIIIIAPSYSEEALAILTNKKKNLRILELPFDAQA  
 Sbjct: 301 SIFGGIIVLNREVDAATAKMHPIFLEIIIIAPSYSEEALAILTNKKKNLRILELPFDAQA 360  
 Query: 361 ASEVEAEYTGTVGGLLVQNQDVVAENPSDWQVVTDRQPTQEATALEFAWKAIKYVKSNG 420  
 ASEVEAEYTGTVGGLLVQNQDVVAENPSDWQVVTDRQPTQEATALEFAWKAIKYVKSNG  
 Sbjct: 361 ASEVEAEYTGTVGGLLVQNQDVVAENPSDWQVVTDRQPTQEATALEFAWKAIKYVKSNG 420  
 Query: 421 IIITNDHMTLGLGAGQTNRVGSVKIAIEQAKDHLDGAVLASDAFFFPADNIEEIAAAGIK 480  
 IIITNDHMTLGLGAGQTNRVGSVKIAIEQAKDHLDGAVLASDAFFFPADNIEEIAAAGIK  
 Sbjct: 421 IIITNDHMTLGLGAGQTNRVGSVKIAIEQAKDHLDGAVLASDAFFFPADNIEEIAAAGIK 480  
 Query: 481 AIIQPGGSVRDQESIDAANKHGLTMIFTGVRHFRH 515  
 AIIQPGGSVRDQ+SIDAANKHGLTMIFTGVRHFRH  
 Sbjct: 481 AIIQPGGSVRDQDSIDAANKHGLTMIFTGVRHFRH 515

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 298

A DNA sequence (GBSx0326) was identified in *S.agalactiae* <SEQ ID 961> which encodes the amino acid sequence <SEQ ID 962>. This protein is predicted to be similar to antibiotic resistance protein. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1842(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12342 GB:Z99106 similar to antibiotic resistance protein

[Bacillus subtilis]

Identities = 65/263 (24%), Positives = 117/263 (43%), Gaps = 34/263 (12%)

Query: 5 KNLEIVESIFGD-WDETIWSCV-QGIMGEVFDSDLPKSSSLAKLGRKSSFGFLAGQPT 62

K ++++F D + T ++S + Q I G V+ D PKS +G +S F+AG

Sbjct: 10 KKYSSLKTMFDDKYCPTFVYSILDQTIPGAVYADDQTFPKSFF--IGTESGIYFIAGDQG 67

Query: 63 -----LFLLEVCSGEDIILVPQHKGWSDLIESTYGCNAHSFKRYATKKDTLERS 112

+ +V S + L W +++ + + +R A +

Sbjct: 68 NRDFHDFIAGYEEQVKSFRFTLFSSSDTWDSVLKPIKDDLQMRRAAFSY-----QP 122

Query: 113 RLEKFVTVQLPNGFELRAIDEKV-----YNSCLEKEWSQDLVANYATYQYVKKQIGYVV 166

+ K QLP G L+ IDE + +NS +E+ + + + +G G+ V

Sbjct: 123 KSFKKTLQLPKGLVLRIDEDIISHSTAFNSAYYEEY-----WNSVSQFASKGFGFAV 175

-382-

Query: 167 YYQGNIIAGASSYSTYKNGIEIEVDTHPDFRRRGLATIVAAQLILTCLDKGIYPSWDAH- 225  
 + +++++ +S N E+++ T ++R GLA VA + I C++ GI PSWD  
 Sbjct: 176 LHGNHVVSECTSI FLGHNRAEMDIYTL EEYRGLGLAYCVANRFIAFCMENGIVPSWD CDI 235

Query: 226 -TRTSLNLSEKLG YEF SHEYIAY 247  
 +S+ L+ KLG++ EY Y  
 Sbjct: 236 CNNSSIALAAKLGFKTVTEYTIY 258

10 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 299

15 A DNA sequence (GBSx0328) was identified in *S.agalactiae* <SEQ ID 963> which encodes the amino acid sequence <SEQ ID 964>. This protein is predicted to be phosphoribosylglycinamide formyltransferase homolog (purN). Analysis of this protein sequence reveals the following:

Possible site: 48  
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.0736(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 965> which encodes the amino acid sequence <SEQ ID 966>. Analysis of this protein sequence reveals the following:

Possible site: 48  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.53 Transmembrane 75 - 91 ( 75 - 91)

30 ----- Final Results -----  
           bacterial membrane --- Certainty=0.1213(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the databases:

>GP:CAA04374 GB:AJ000883 purD [Lactococcus lactis]  
 Identities = 236/419 (56%), Positives = 301/419 (71%), Gaps = 7/419 (1%)

40 Query: 50 LKLLVVGSGGREHAIAKLLASKGVDQVFVAPGNDGMTLDGLDLVNIIVVSEHSRLIAFAK 109  
 +K+LV+GSGGREHA+AKK + S V++VFVAPGN GM DG+ +V+I + +L+ FA+  
 Sbjct: 1 MKILVIGSGGREHALAKKFMESPQVEEVFVAPGNSGMEKDGIQIVHISELSNDKLVKFAQ 60

45 Query: 110 ENEISWAFIGPDDALAAGIVDDFNSAGLRAFPGPTKAAAELEWSKDFAKEIMVKYNVPTAA 169  
 I F+GP+ AL G+VD F A L FGP K AAELE SKDFAK IM KY VPTA  
 Sbjct: 61 NQNIGLTFVGPETALMNGVVDAFIKAE LPIFGPNKMAAELEGSKDFAKSIMKKYGVPTAD 120

50 Query: 170 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 229  
 Y TF E A AY++E+G P+V+KADGLA GKG V VA +E A A ++ F S  
 Sbjct: 121 YATFDSLEPALAYLDEKGVPLVIKADGLAAGKGVTVAFDIETAKSALADI-----FSGSQ 175

55 Query: 230 ARVVIEEFLDGE EFSLF AFANGDKFYIMPTAQDHKRAFDGDKGPNTGGMGAYAPVPHLPQ 289  
 +VVIEEFLDGE EFSLF +F + K Y MP AQDHKRAFD DKGPNTGGMGAY+PV H+ +  
 Sbjct: 176 GKVVIEEFLDGE EFSLFSFIHDGKIYPMPIAQDHKRAFDGDKGPNTGGMGAYSPVLHISK 235

Query: 290 SVVDTAVEMIVRPVLEG MVAEGRPYLGVLYVGLILTADGPKVIEFN SRFGDPETQIILPR 349  
 VV+ A+E +V+P + GM+ EG+ + GVLY GLILT DG K IEFN+RFGDPETQ++LPR  
 Sbjct: 236 EVVNEALEKVVKPTVAGMIEEGKSFTGVLYAGLILTEDGVKTIEFNARFGDPETQVVLPR 295



-383-

Query: 350 LTSDFQAQNIDDIMMGIEPYITWQKDGVTLGVVVASEGYFPDYEKGVPLPEKTDGDIITYY 409  
 L SD AQ I DI+ G EP + W + GVTLGVVVA+EGYP + G+ LPE +G + YY  
 Sbjct: 296 LKSDLAQAIIDILAGNEPTLEWLESGVTLGVVVAAGYPSQAKLGLILPEIPEG-LNVYY 354

Query: 410 AGVKFSENSELILLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQDQDTTGLFYRNDIGSKAI 468  
 AGV +EN++ L+S+GGRVY++ T + VK+ Q +Y +L + + G FYR+DIGS+AI  
 Sbjct: 355 AGVSKNENNQ-LISSGGRVYLVSETGEDVKSTQKLLYEKLDKLENDGFFYRHDIGSRAI 412

10 An alignment of the GAS and GBS proteins is shown below:

Identities = 172/182 (94%), Positives = 176/182 (96%)

Query: 1 MKIAVFASGNGSNFQVIAEQFQVSFVFSHDRDAYVLERAQNLAIPSFAPFELKEFENKAAY 60  
 MKIAVFASGNGSNFQVIAEQF VSFVFSHDRDAYVLERAQNLAIPSFAPFELKEFENK AY  
 Sbjct: 1 MKIAVFASGNGSNFQVIAEQFQVSFVFSHDRDAYVLERAQNLAIPSFAPFELKEFENKVAAY 60

Query: 61 EQAVVDLLDKHEIDLVLCLAGYMKIVGETLLSAYEGRIINIHPYLPFPGAHGIKDAWEA 120  
 EQA+VDLLDKHEIDLVLCLAGYMKIVGETLL AYE RIINIHP YLPFPGAHGI+DAWEA  
 Sbjct: 61 EQAIVDLLDKHEIDLVLCLAGYMKIVGETLLLAYERRIINIHPAYLPFPGAHGIEDAWEA 120

Query: 121 GVDQSGVTIHWVDSGVDTGQVIQVHVPRLADDSLESFETRIHETEQLYPAVLDSLGIK 180  
 GVDQSGVTIHWVDSGVDTGQVIQV VPRLADDSLESFETRIHETEQLYPAVLDSLGI++  
 Sbjct: 121 GVDQSGVTIHWVDSGVDTGQVIQVVRPRLADDSLESFETRIHETEQLYPAVLDSLGE 180

Query: 181 RK 182  
 RK  
 Sbjct: 181 RK 182

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 vaccines or diagnostics.

### Example 300

A DNA sequence (GBSx0329) was identified in *S.agalactiae* <SEQ ID 967> which encodes the amino acid  
 sequence <SEQ ID 968>. Analysis of this protein sequence reveals the following:

Possible site: 52  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.59 Transmembrane 121 - 137 ( 121 - 137)

----- Final Results -----  
 bacterial membrane --- Certainty=0.1235(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC16901 GB:AF016634 phosphoribosylformylglycinamide  
 cyclo-ligase [Lactococcus lactis subsp. cremoris]  
 Identities = 253/338 (74%), Positives = 288/338 (84%), Gaps = 4/338 (1%)

Query: 4 KNAYAQSGVDVEAGYEVVERIKKHVARTERAGVMGALGGFGGMFDLSQTGVKEPVLISGT 63  
 +NAYA+SGVDVEAGYEVV RIKKHVA+TER GV+GALGGFGG FDLS VKEPVLISGT  
 Sbjct: 5 ENAYAKSGVDVEAGYEVVSRIKKHVAKTERLGVLGALGGFGGSFDLSVLDVKEPVLISGT 64

Query: 64 DGVGTKMLAIKYDKHDTIGQDCVAMCVNDIIAAGAEPLYFLDYVATGKNEPAKLEQVVA 123  
 DGVGTKMLAI+ DKHDTIG DCVAMCVNDIIAAGAEPLYFLDY+ATGKN P KLEQVVA  
 Sbjct: 65 DGVGTKMLAIRADKHDTIGIDCVAMCVNDIIAAGAEPLYFLDYIATGKNIPEKLEQVVA 124

Query: 124 GVAEGCVQASALIGGETAEMPGMYGEDDYDLAGFAVGVAEKSQIIDGSK-VKEGDILLG 182  
 GVAEGC+QA AALIGGETAEMPGMY EDDYDLAGFAVGVAEKSQ+IDG K V+ GD+LLG  
 Sbjct: 125 GVAEGCLQAGAALIGGETAEMPGMYDEDDYDLAGFAVGVAEKSQILIDGEKDV EAGDVLLG 184

Query: 183 LASSGIHSNGYSLVRRVFADYTGDEVLPLEGGKQLKDVLLLEPTRIYVKAALPLIKEELVN 242  
 LASSGIHSNGYSLVR+VFAD+ +E LPEL+ + L D LL PT+IYVK LPLIK+ +

-384-

Sbjct: 185 LASSGIHSNGYSLVRKVFADFNLNLSLPELD-QSLIDTLLTPTKIYVKELLPLIKQNKIK 243

Query: 243 GIAHITGGGFIENVPRMFADDLAAEIDEDKVPVLPIFKALEKYGDIKHEEMFEIFNMVG 302  
 GIAHITGGGF EN+PRMF + L+AEI E VLPPIFKALEKYG IKHEEM+EIFNMG+G

5 Sbjct: 244 GIAHITGGGFHENLPRMFGNSLSAEIVEGSDVLPPIFKALEKYGSIKHEEMYEIFNMGIG 303

Query: 303 LMLDVNPENVDVRVKELLDEPVYEIGRIIKKADDSVVIK 340  
 +++ V PEN +K+ L+ +EIG+++ + + VVIK

10 Sbjct: 304 MVIAPENAAALKKELN--AFEIGQMVNRQAPVVIK 339

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 969> which encodes the amino acid sequence <SEQ ID 970>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3236(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 321/340 (94%), Positives = 332/340 (97%)

25 Query: 1 MSEKNAYAQSGVDVEAGYEVVERIKKHVARTERAGVMGALGGFGGMFDLSQTGVKEPVL 60  
 MSEKNAYA+SGVDVEAGYEVVERIKKHVARTERAGVMGALGGFGGMFDLS+TGVKEPVL+

Sbjct: 1 MSEKNAYAKSGVDVEAGYEVVERIKKHVARTERAGVMGALGGFGGMFDLSKTGVKEPVLV 60

30 Query: 61 SGT DGVG TKLM LAIKYDKHDTIGQDCVAMCVNDIIAAGAEPLYFLDYVATGKNPAKLEQ 120  
 SGT DGVG TKLM LAIKYDKHDTIGQDCVAMCVNDIIAAGAEPLYFLDY+ATGKN P KLE+

Sbjct: 61 SGT DGVG TKLM LAIKYDKHDTIGQDCVAMCVNDIIAAGAEPLYFLDYIATGKNNPVKLEE 120

35 Query: 121 VVAGVAEGCVQASAALIGGETAEMPGMYGEDDYDLAGFAVGVAEKSQIIDGSKVKEGDIL 180  
 VV+GVAEGCVQA AALIGGETAEMPGMYG+DDYDLAGFAVGVAEKSQIIDGSKVKEGDIL

Sbjct: 121 VVSGVAEGCVQAGAALIGGETAEMPGMYGQDDYDLAGFAVGVAEKSQIIDGSKVKEGDIL 180

40 Query: 181 LGLASSGIHSNGYSLVRRVFADYTGDEVLPLELEGKQLKDVLLPETHRIYVKAALPLIKEEL 240  
 LGLASSGIHSNGYSLVRRVFADYTG E+LPELEGKQLKDVLLPETHRIYVKAALPLIKEEL

Sbjct: 181 LGLASSGIHSNGYSLVRRVFADYTGKELLPELEGKQLKDVLLPETHRIYVKAALPLIKEEL 240

45 Query: 241 VNGIAHITGGGFIENVPRMFADDLAAEIDEDKVPVLPIFKALEKYGDIKHEEMFEIFNMG 300  
 V GI HITGGGFIEN+PRMFADDLAAEIDEDKVPVLPIFKALEKYGDIKHEEMFEIFNMG

Sbjct: 241 VKGIGHITGGGFIENIPRMFADDLAAEIDEDKVPVLPIFKALEKYGDIKHEEMFEIFNMG 300

Query: 301 VGLMLDVNPENVDVRVKELLDEPVYEIGRIIKKADDSVVIK 340  
 VGLML V+PENVRVKELLDEPVYEIGRIIKKAD SVVIK

Sbjct: 301 VGLMLAVSPENVRVKELLDEPVYEIGRIIKKADASVVIK 340

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 301

A DNA sequence (GBSx0330) was identified in *S.agalactiae* <SEQ ID 971> which encodes the amino acid sequence <SEQ ID 972>. This protein is predicted to be phosphoribosylpyrophosphate amidotransferase (purF). Analysis of this protein sequence reveals the following:

Possible site: 25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1112(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-385-

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database:

5 >GP:AAD12627 GB:U64311 phosphoribosylpyrophosphate amidotransferase  
[Lactococcus lactis]  
Identities = 340/470 (72%), Positives = 404/470 (85%), Gaps = 6/470 (1%)

10 Query: 3 YE VKSLNEECGVFGI WGY PQA AQVTY FGLHSLQHRGQEGAGIISNDNGKLYGYRNVGLLS 62  
+E K+LNEECG+FG+WG+P AA++TYFGLH+LQHRGQEGAGI+ N+NGKL +R +GL++  
Sbjct: 37 FEAKTLNEECGLFGVWGHPDAARLT YFGLHALQHRGQEGAGILVNNNGKLNRRHRLGLVT 96

15 Query: 63 EVFKNQSELDNLTGNA AIGHVRYATAGSADIRNIQPF LYKFHDGQFALCHNGNLTNAISS 122  
EVF+++ +L+ LTG++AIGHVRYATAGSA+I NIQPF ++FHDG L HNGNLTNA S  
Sbjct: 97 EVFRHEKDLEELTGSSAIGHVRYATAGSANINNIQPFQFEFHDGSLGLAHNGNLTNAQSL 156

20 Query: 123 RKELEKQGAIFNASSDTEILMHLIRSHNPSFMGKVKEALSTVKGGFAYLLMTEDKLI AA 182  
R ELEK GAIF+++SDTEILMHLIRSH+P FMG+VKEAL+TVKGGFAYL+MTE+ +AAA  
Sbjct: 157 RCELEKSGAIFSSNSDTEILMHLIRSHHPEFMGRVKEALNTVKGGFAYLIMTENSIVAA 216

25 Query: 183 LDPNFRPLSIGQMONGAWVISSETCAFEVVGAKWVRDVEPGEVILIDDSGIQC DRYTDE 242  
LDPN FRPLSIG+M NGA V++SETCAF+VVG A W++DV+PGE+I I+D GI D++TD  
Sbjct: 217 LDPNGFRPLSIGKMSGALVVA SETCAF DVVGATW IQDVQPGEIIEINDDGIHVDQFTDS 276

30 Query: 243 TQLAICSM EYVYFARPDSTIHGVNVHTARKNMGKRLAQEFKQDADIVIGVPNSSLSAAMG 302  
T + ICSMEY+YFARPDS I GVNVTARK GK LAQE K DADIVIGVPNSSLSAA G  
Sbjct: 277 TNMTICSM EYIYFARPDSNIAGVNVHTARKRSGKILAQEAKIDADIVIGVPNSSLSAASG 336

35 Query: 303 FAESGLPNEMGLVKNQYRTFTIQPTQELREQGVRMKLSAVSGVVGKRVVMIDDSIVR 362  
+AESGLP EMGL+KNQY RTFTIQPTQELREQGVRMKLSAV GVV+GKRV+M+DDSI VR  
Sbjct: 337 YAESGLPYEMGLIKNQYVARTFTIQPTQELREQGVRMKLSAVRGVVEGKRVIMVDDSI VR 396

Query: 363 GTTSRRIVGLLREAGATEVHVAIASPELKYPFCFYGIDIQTRRELISANHAVDEVCDIIGA 422  
GTTSRRIV LL++AGA EVHVAIASP LKYPFCFYGIDIQ R ELI+A H DE+ + IGA  
Sbjct: 397 GTTSRRIVKLLKDAGAAEVHVAIASPALKYPFCFYGIDIQDRDELIAATHTTTDEIREAIGA 456

40 Query: 423 DSLTYLSIDGLIKSIGLETKAPNGGLCVAYFDGHYPTPLYDYEEYLRSL 472  
DSLTYLS GL+++IG + LC++YFDG YPTPLYDYE +YL SL  
Sbjct: 457 DSLTYLSQSGLEVAIG-----HDKLCLSYFDGEYPTPLYDYEADYLESL 500

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 973> which encodes the amino acid sequence <SEQ ID 974>. Analysis of this protein sequence reveals the following:

Possible site: 21  
>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.0610 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50 An alignment of the GAS and GBS proteins is shown below:

Identities = 473/484 (97%), Positives = 481/484 (98%)

55 Query: 1 MTYEVKSLNEECGVFGI WGY PQA AQVTY FGLHSLQHRGQEGAGIISNDNGKLYGYRNVGL 60  
MTYEVKSLNEECGVFGI WGY PQA AQVTY FGLHSLQHRGQEGAGI+SNNDNGKLYGYRNVGL  
Sbjct: 20 MTYEVKSLNEECGVFGI WGY PQA AQVTY FGLHSLQHRGQEGAGIVSNNDNGKLYGYRNVGL 79

60 Query: 61 LSEVFKNQSELDNLTGNA AIGHVRYATAGSADIRNIQPF LYKFHDGQFALCHNGNLTNAI 120  
LSEVFKNQSELDNLTGNA AIGHVRYATAGSADIRNIQPF LYKFHDGQFALCHNGNLTNAI  
Sbjct: 80 LSEVFKNQSELDNLTGNA AIGHVRYATAGSADIRNIQPF LYKFHDGQFALCHNGNLTNAI 139

Query: 121 SSRKELEKQGAIFNASSDTEILMHLIRSHNPSFMGKVKEALSTVKGGFAYLLMTEDKLI 180  
S RKELEKQGAIFNASSDTEILMHLIRSHN SFMGKVKEAL+TVKGGFAYLLMTE+KLI  
Sbjct: 140 SLRKELEKQGAIFNASSDTEILMHLIRSHNSSFMGKVKEALNTVKGGFAYLLMTENKLI 199

-386-

Query: 181 AALDPNAFRPLSIGQMONGAWVISSETCAFEVVGAKWVRDVEPGEVILIDDSGIQCDRYT 240  
 AALDPNAFRPLSIGQMONGAWVISSETCAFEVVGAKWVRDVEPGEVILIDD GIQCDRYT  
 Sbjct: 200 AALDPNAFRPLSIGQMONGAWVISSETCAFEVVGAKWVRDVEPGEVILIDDRGIQCDRYT 259

5 Query: 241 DETQLAICSMEEVYFARPDSTIHGVNVHTARKNMGKRLAQEFKQDADIVIGVPNSSLSAA 300  
 DETQLAICSMEEVYFARPDSTIHGVNVHTARKNMGKRLAQEFKQDADIVIGVPNSSLSAA  
 Sbjct: 260 DETQLAICSMEEVYFARPDSTIHGVNVHTARKNMGKRLAQEFKQDADIVIGVPNSSLSAA 319

10 Query: 301 MGFAEESGLPNEMGLVKNQYTQRTFIQPTQELREQGVRMKLSAVSGVVKGKRVVMIDDSI 360  
 MGFAEESGLPNEMGLVKNQYTQRTFIQPTQELREQGVRMKLSAVSGVVKGKRVVMIDDSI  
 Sbjct: 320 MGFAEESGLPNEMGLVKNQYTQRTFIQPTQELREQGVRMKLSAVSGVVKGKRVVMIDDSI 379

15 Query: 361 VRGTTSRRIVGLLREAGATEVHVAIASPELKYPFCFYGIDIQTRRELISANHAVDEVCDII 420  
 VRGTTSRRIVGLLREAGA+EVHVAIASPELKYPFCFYGIDIQTRRELISANH+VDEVCDII  
 Sbjct: 380 VRGTTSRRIVGLLREAGASEVHVAIASPELKYPFCFYGIDIQTRRELISANHSVDEVCDII 439

20 Query: 421 GADSLTYLSIDGLIKSIGLETKAPNGGLCVAYFDGHYPTPLYDYEEYLRSLLEEKTSFYI 480  
 GADSLTYLS+DGLI+SIGLETKAPNGGLCVAYFDGHYPTPLYDYEEYLRSLLEEKTSFYI  
 Sbjct: 440 GADSLTYLSLDGLIESIGLETKAPNGGLCVAYFDGHYPTPLYDYEEYLRSLLEEKTSFYI 499

Query: 481 QKVK 484  
 QKVK  
 Sbjct: 500 QKVK 503

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 302

A DNA sequence (GBSx0331) was identified in *S.agalactiae* <SEQ ID 975> which encodes the amino acid sequence <SEQ ID 976>. Analysis of this protein sequence reveals the following:

30 Possible site: 28  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 35 bacterial cytoplasm --- Certainty=0.4797(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 303

A DNA sequence (GBSx0332) was identified in *S.agalactiae* <SEQ ID 977> which encodes the amino acid sequence <SEQ ID 978>. Analysis of this protein sequence reveals the following:

45 Possible site: 13  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 50 bacterial cytoplasm --- Certainty=0.3489(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-387-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 304

A DNA sequence (GBSx0333) was identified in *S.agalactiae* <SEQ ID 979> which encodes the amino acid sequence <SEQ ID 980>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10                   bacterial cytoplasm --- Certainty=0.1690(Affirmative) < succ>  
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15           >GP:CAC12194 GB:AL445066 phosphoribosylformylglycinamide synthase  
            related protein [Thermoplasma acidophilum]  
            Identities = 199/746 (26%), Positives = 329/746 (43%), Gaps = 103/746 (13%)

20           Query: 202 ADD--FAAYKAEQGLAMEVDDLFIQDYFKSIGRVPTETELKVLDTYWSHCRHTTFETE 259  
            ADD    A    GLA+ +D++ ++ YF+ +GR P + E+ +   WS+HC + + +  
            Sbjct: 11 ADDARLKAISKRLGLALSLEDEMKAVRSYFERLGRDPIDAEIHAVAQSWSEHCSYKSSKYY 70

            Query: 260 LKNIDFSASKFQKQLQATYDKYIAMRDELGRSEKPTLMDMATIFGRYERANGRLDDMEV 319  
                      LK       K+   L+   Y   +AM D+ G  
25           Sbjct: 71 LK-----KYLGSCLKTDYT-ILAMEDDAG----- 92

            Query: 320 SDEINACSVEIEVDVDGVKEPWLMLFKNETHNHPTETIEPFGGAATCIGGAIRDPLSGRSY 379  
                          VD DG   + + K E+HNHP+ +EP+GGAAT IGG +RD L   +  
30           Sbjct: 93 -----VVDFDG---EYAYVLKMESHNHPSAVEPYGGAATGIGGIVRDVLCMGAQ 138

            Query: 380 VYQAMRISGAGDITTPIAETRAGKLPQQVISKTAAGHYSSYGNQIGLATTYVREYFHPGF 439  
                      +       GD+++ E   G L + I       G   YGN+IG+       YF +  
            Sbjct: 139 PVALIDSLFLGDVSSDRYE---GLLSPRYIFGGVVGIGIRDYGNRIGIPNVAGSLYFDKLY 195

35           Query: 440 VAKRMELGAVVGAAPKENNVREKP-EAGDVVLLGGKTGRDGVGGATGSSKVTVESVET 498  
                      + +       VG   ++ +VR K + GDV+VL+GGKTGRDG+ G   +S   + ++  
            Sbjct: 196 NSNPLVNAGCVGIVRRDRIVRSKSYKPGDVLVLMGGKTGRDGIHGVNFASTTLG-KVTKS 254

            Query: 499 AGAEVQKGNAIERKIQRFLRDGNVTRLIKKSNDFGAGGVCVAIGELAD---GLEIDL 554  
                      +   +Q GN I E+ + + + N   LI+   D G GG+ A E+       G EI LD  
40           Sbjct: 255 SRLAIQLGNPIVEQPMIKAVLEANDAGLIRAMKDLGGGGLSSAATEMVYAGGFGAEITLD 314

            Query: 555 KVPLKYQGLNGTEIAISESQERMSVVVGPSDVDAFIAACNKENIDAVVATVTEKPNLVM 614  
                      + LK   ++G EI ISESQERM +   P DV+       K N+D V+ VT   + +  
45           Sbjct: 315 DIKLKESNMGSWEIWISESQERMLMECYPEDVEKIRQIAEKWNLDFSVIGQVTADRRIRV 374

            Query: 615 TWNGETIVDLERCFLDTNGV-RVVVDKVVVDKDLTVPEARNTSAETLEADMLKVLSDLNH 673  
                      +   I+D++ FLD + V +       K V+K +TVP+       E L + +   ++ LN  
50           Sbjct: 375 YYKKRKIIDMDIEFLDDSPVYQRPYRIKEVEKSVTVPQ----EPEDLNSFVRDFMARLNT 430

            Query: 674 ASQKGLQTIFDSSVGRSTVNHPIGGR-YQITPTESSVQKLPVQYGVTTTASVMAQGYNPY 732  
                      ++ +   +D +V ST+ P GR + T +++V K P++ +       V+ G P  
            Sbjct: 431 CARFNVVRQYDHTVVGSTIVTPFVGRPNKETHADATVIK-PLNSM--RGLVLTSGSRPN 487

55           Query: 733 IAEWSPYHGAAYAVIEATARLVATGADWSRARFSYQYFERMDKQAEFRGQPVSAALLGSI 792  
                      +   PY G   + EA   +++TG   R               ++ E GQ V ++  
            Sbjct: 488 MVSVDPAVAGTLLTLAEAYKNILSTG---GRPHSVVDALNFGNPEREIMGQFVESVRAIG 544

            Query: 793 EAQIQFGLPSIGGKDSMSGTFEELTVPPTLVAFGVTTADS-RKVLSPFEKAAAGENIY--- 848  
                      +   + GLP + G S   + + + PT       V D R+ +   K +G IY  
60           Sbjct: 545 DFCRKMGLPVVAGNVSFYNEYRKTDIMPTPTIMMVGLIDDVRRSRTTYMKSGSNAIYLIG 604

-388-

Query: 849 -----YIPGQAISEIDIDFLIKANF--SQFEAIQAQHKITAASAVKYGG 890  
                                   Y  G  + D+D          +F S+ + I + H +++          GG  
 Sbjct: 605 EPCDNLGTGSEYSRMHGTYTDGFLPAPDLDELTRIRDFLSSKADMILSSHVDVSS-----GG 658

5       Query: 891 VLESALMTFGNRRIGASVEIAELDSS 916  
                                   +  +L+ M+FG+ IG  V+I+ + ++  
 Sbjct: 659 LFAALSEMSFGSGIGFHVDISNVSA 684

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 981> which encodes the amino acid  
 10 sequence <SEQ ID 982>. Analysis of this protein sequence reveals the following:

Possible site: 51  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15               bacterial cytoplasm --- Certainty=0.1415(Affirmative) < succ>  
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

20       Identities = 1219/1256 (97%), Positives = 1226/1256 (97%)

Query: 11   SSYFRVAPLSDLVSYMNKRIFVEKKADFGIKSASLVKELTHNLQLASLKDLRIVQVYDVF 70  
           SSYF  VAPLSDLVSYMNKRIFVEKKADFGIKSASLVKELTHNLQL  SLK  LRIVQVYDVF  
 Sbjct: 2   SSYFPVAPLSDLVSYMNKRIFVEKKADFGIKSASLVKELTHNLQLTSLKALRIVQVYDVF 61

25       Query: 71   NLAEDLLARAEEKHIFSEQVTDRLLEAEITAEIDKVAFFAIEALPGQFDQRAASSQEALL 130  
           NLAEDLLARAEEKHIFSEQVTD  LLTE  EITAEIDKVAFFAIEALPGQFDQRAASSQEALL  
 Sbjct: 62   NLAEDLLARAEEKHIFSEQVTDCLLTETEITAEIDKVAFFAIEALPGQFDQRAASSQEALL 121

30       Query: 131   LLGSDSQVKVNTAQLYLVNKDIAEAELEAVKNYLLNPVDSRFKDITLPLEVQAFSVSDKT 190  
           L  GSDSQVKVNTAQLYLVNKDI  EAELEAVKNYLLNPVDSRFKDITLPLE  QAFSVSDKT  
 Sbjct: 122   LFGSDSQVKVNTAQLYLVNKDITEAELEAVKNYLLNPVDSRFKDITLPLEEQAFSVSDKT 181

35       Query: 191   ISNLDFFETYQADDFAAAYKAEQGLAMEVDDLFIQDYFKSIGRVPTETELKVLDTYWSDH 250  
           I  NLDFFETYQADDF  YKAEQGLAMEVDDLFIQ+YFKSIG  VPTETELKVLDTYWSDH  
 Sbjct: 182   IPNLDFFETYQADDFATYKAEQGLAMEVDDLFIQNYFKSIGCVPTETELKVLDTYWSDH 241

40       Query: 251   CRHTTFETELKNIDFSASKFQKQLQATYDKYIAMRDELGRSEKPTLMDMATIFGRYERA 310  
           CRHTTFETELKNIDFSASKFQKQLQ  TYDKYIAMRDELGRSEKPTLMDMATIFGRYERA  
 Sbjct: 242   CRHTTFETELKNIDFSASKFQKQLQTTYDKYIAMRDELGRSEKPTLMDMATIFGRYERA 301

45       Query: 311   NGRLDDMEVSDEINACSVEIEVDVDGVKEPWLLMFKNETHNHPTEIEPFPGAATCIGGAI 370  
           NGRLDDMEVSDEINACSVEIEVDVDGVKEPWLLMFKNETHNHPTEIEPFPGAATCIGGAI  
 Sbjct: 302   NGRLDDMEVSDEINACSVEIEVDVDGVKEPWLLMFKNETHNHPTEIEPFPGAATCIGGAI 361

50       Query: 371   RDPLSGRSYVYQAMRISGAGDITTPIAETRAGKLPQQVISKTAAHGYSSYGNQIGLATTY 430  
           RDPLSGRSYVYQAMRISGAGDITTPIAETRAGKLPQQVISKTAAHGYSSYGNQIGLATTY  
 Sbjct: 362   RDPLSGRSYVYQAMRISGAGDITTPIAETRAGKLPQQVISKTAAHGYSSYGNQIGLATTY 421

55       Query: 431   VREYFHPGFVAKRMELGAVVGAAPKENVVREKPEAGDVVVLLGGKTGRDGVGGATGSSKV 490  
           VREYFHPGFVAKRMELGAVVGAAPKENVVREKPEAGDVV+LLGGKTGRDGVGGATGSSKV  
 Sbjct: 422   VREYFHPGFVAKRMELGAVVGAAPKENVVREKPEAGDVVILLGGKTGRDGVGGATGSSKV 481

60       Query: 491   QTVESVETAGAEVQKGNAIIEERKIQRLFRDGNVTRLIKKSNDFGAGGVCVAIGELADGLE 550  
           QTVESVETAGAEVQKGNAIIEERKIQRLFRDGNVTRLIKKSNDFGAGGVCVAIGELADGLE  
 Sbjct: 482   QTVESVETAGAEVQKGNAIIEERKIQRLFRDGNVTRLIKKSNDFGAGGVCVAIGELADGLE 541

65       Query: 551   IDLDKVPLKYQGLNGTEIAISESQERMSVVVGPSDVFDAFIAACNKENIDAVVVATVTEKP 610  
           IDLDKVPLKYQGLNGTEIAISESQERMSVVV  P+DVDAFIAACNKENIDAVVVATVTEKP  
 Sbjct: 542   IDLDKVPLKYQGLNGTEIAISESQERMSVVVRPNDVFDAFIAACNKENIDAVVVATVTEKP 601

      Query: 611   NLVMTWNGEITVDLERCFLDTNGVRVVVDKAVVDKDLTVPEARTTSAETLEADMLKVLSD 670  
           NLVMTWNGE  IVDLER  FLDTNGVRVVVDKAVVDKDLTVPEARTTSAETLEAD  LKVLSD  
 Sbjct: 602   NLVMTWNGEIIVDLERRFLDTNGVRVVVDKAVVDKDLTVPEARTTSAETLEADTLKVLSD 661

-389-

Query: 671 LNHASQKGLQTI F DSSVGRSTVNHP IGGRYQITPTESSVQKLPVQYGVTTTASVMAQGYN 730  
 LNHASQKGLQTI F DSSVGRSTVNHP IGGRYQITPTESSVQKLPVQ+GVTTTASVMAQGYN  
 Sbjct: 662 LNHASQKGLQTI F DSSVGRSTVNHP IGGRYQITPTESSVQKLPVQHGVTTTASVMAQGYN 721

5 Query: 731 PYIAEWSPYHGAAYAVIEATARLVATGADWSRARFSYQYEFERMDKQAEFRGQPVSAALLG 790  
 PYIAEWSPYHGAAYAVIEATARLVATGADWSRARFSYQYEFERMDKQAEFRGQPVSAALLG  
 Sbjct: 722 PYIAEWSPYHGAAYAVIEATARLVATGADWSRARFSYQYEFERMDKQAEFRGQPVSAALLG 781

10 Query: 791 SIEAQIQFGLPSIGGKDSMSGTFEELTVPPTLVAFGVTTADSRKVLSPFEKAAGENIYYI 850  
 SIEAQIQ GLPSIGGKDSMSGTFE+LTVPPTLVAFGVTTADSRKVLSPFEKAAGENIYYI  
 Sbjct: 782 SIEAQIQLGLPSIGGKDSMSGTFEDLTVPPTLVAFGVTTADSRKVLSPFEKAAGENIYYI 841

15 Query: 851 PGQAISEDIDFDLIKANFSQFEAIQAQHKITAASAVKYGGVLESALMTFGNRIGASVEI 910  
 PGQAISEDIDFDLIK NFSQFEAIQAQHKITAASA KYGGVLESALMTFGNRIGASVEI  
 Sbjct: 842 PGQAISEDIDFDLIKDNFSQFEAIQAQHKITAASAAKYGGVLESALMTFGNRIGASVEI 901

20 Query: 911 AELDSSLTAQLGGFVFTSVEEIIADVVKIGQTQADFTVTVNGNDLAGASLLSAFEGKLEEV 970  
 AELDSSLTAQLGGFVFTS EEEIAD VKIGQTQADFTVTVNGNDLAGASLL+AFEGKLEEV  
 Sbjct: 902 AELDSSLTAQLGGFVFTSABEEIADAVKIGQTQADFTVTVNGNDLAGASLLAAAFEGKLEEV 961

25 Query: 971 YPTEFEQVDAIEEVPAAVSDVVIIKAKEIEKPVVYIPVFPGTNSEYDSAKAFEQVGASVN 1030  
 YPTEFEQ D +EEVPAVSD VIKAKE IEKPVVYIPVFPGTNSEYDSAKAFEQVGASVN  
 Sbjct: 962 YPTEFEQTDVLEEVPAAVSDTVIKAKETIEKPVVYIPVFPGTNSEYDSAKAFEQVGASVN 1021

30 Query: 1031 LVPPFVTLNEAAIAESVDTMVANIAKANIIFFAGGFSAADEPDGSAKFIVNILLNEKVRAA 1090  
 LVPPFVTLNE AIAESVDTMVANIAKANIIFFAGGFSAADEPDGSAKFIVNILLNEKVRAA  
 Sbjct: 1022 LVPPFVTLNEVAIAESVDTMVANIAKANIIFFAGGFSAADEPDGSAKFIVNILLNEKVRAA 1081

35 Query: 1091 IDSFIEKGGLIIGICNGFQALVKSGLLPYGNFEEAGETSPTLFYNDANQHVAKMVETRIA 1150  
 IDSFIEKGGLIIGICNGFQALVKSGLLPYGNFEEAGETSPTLFYNDANQHVAKMVETRIA  
 Sbjct: 1082 IDSFIEKGGLIIGICNGFQALVKSGLLPYGNFEEAGETSPTLFYNDANQHVAKMVETRIA 1141

40 Query: 1151 NTNSPWLAGEVEVDIHVIPVSHGEGKFVVSASEFAELRDNGQIWSQYVDFDQPSMDSKY 1210  
 NTNSPWLAGEVEVDIH IPVSHGEGK VVSASEFAELRDNGQIWSQYVDFDQPSMDSKY  
 Sbjct: 1142 NTNSPWLAGEVEVDIHAIPVSHGEGKLVVSASEFAELRDNGQIWSQYVDFDQPSMDSKY 1201

Query: 1211 NPNGSVNAIEGITSKNGQIIGKMGHSERWEDGLFQNI PGNKDQKLFESAVKYFTGK 1266  
 NPNGSVNAIEGITSKNGQIIGKMGHSERWEDGLFQNI PGNKDQ LF SAVKYFTGK  
 Sbjct: 1202 NPNGSVNAIEGITSKNGQIIGKMGHSERWEDGLFQNI PGNKDQILFASAVKYFTGK 1257

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 305

A DNA sequence (GBSx0334) was identified in *S. agalactiae* <SEQ ID 983> which encodes the amino acid sequence <SEQ ID 984>. This protein is predicted to be phosphoribosylaminoimidazole-succinocarboxamide synthase (purC). Analysis of this protein sequence reveals the following:

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence

#### ----- Final Results -----

bacterial cytoplasm --- Certainty=0.4783(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA03540 GB:L15190 SAICAR synthetase [Streptococcus pneumoniae]  
 Identities = 183/231 (79%), Positives = 203/231 (87%)

Query: 1 MTNQLIYTGKAKDIYSTKDENVIRTVYKDQATMLNGARKETIDGKGALNNQISSLIFEKL 60  
 M+ QLIY+GKAKDIY+T+DEN+I + YKDQAT NG +KE I GKG LNNQISS IFELK  
 Sbjct: 1 MSKQLIYSGKAKDIYTTEDENLIISTYKDQATAFNGVKKEQIAGKGVLLNNQISSFIFEKL 60

-390-

Query: 61 NMAGVVTHTYIEQISKNEQLNKKVDIIPLEVLRNVTAGSFSKRFGVEEGHVLETPIVEFY 120  
 N AGV TH++E++S EQLNKKV IIPLEVLRN TAGSFSKRFGV+EG LETPIVEFY  
 5 Sbjet: 61 NAAGVATHFVEKLSDEQLNKKVKIIPLEVLRNYTAGSFSKRFGVDEGIALETPIVEFY 120

Query: 121 YKNDNLNDPFINDEHVKFLGIVNDEEIAYLKGETRHINELLKDWFAQIGLNLIDFKLEFG 180  
 YKND+L+DPPFINDEHVKFL I +D++IAYLK E R INELLK WFA+IGL LIDFKLEFG  
 Sbjet: 121 YKNDLDDPFINDEHVKFLQIADDQIAYLKKEARRINELLKVVWFAEIGLKLIDFKLEFG 180

10 Query: 181 FDKDGKIILADEFSPDNCRLWDADGNHMDKDVFRDLGSLTDVYQVVLEKL 231  
 FDKDGKIILADEFSPDNCRLWDADGNHMDKDVFR LG LTDVY++V EKL  
 Sbjet: 181 FDKDGKIILADEFSPDNCRLWDADGNHMDKDVFRRLGELTDVYEVWEKL 231

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 985> which encodes the amino acid  
 15 sequence <SEQ ID 986>. Analysis of this protein sequence reveals the following:

Possible site: 51  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.3935(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

25 Identities = 221/234 (94%), Positives = 228/234 (96%)

Query: 1 MTNQLIYTGKAKDIYSTKDENVIRTVYKDQATMLNGARKETIDGKGALNNQISSLIFEKL 60  
 +TNQLIY GKAKDIYSTKDENVIRTVYKDQATMLNGARKETIDGKGALNNQISSLIFEKL  
 30 Sbjet: 11 VTNQLIYKGKAKDIYSTKDENVIRTVYKDQATMLNGARKETIDGKGALNNQISSLIFEKL 70

Query: 61 NMAGVVTHTYIEQISKNEQLNKKVDIIPLEVLRNVTAGSFSKRFGVEEGHVLETPIVEFY 120  
 N AGVVTHTYIEQISKNEQLNKKVDIIPLEVLRNVTAGSFSKRFGVEEGHVLETPIVEFY  
 Sbjet: 71 NKAGVVTHTYIEQISKNEQLNKKVDIIPLEVLRNVTAGSFSKRFGVEEGHVLETPIVEFY 130

35 Query: 121 YKNDNLNDPFINDEHVKFLGIVNDEEIAYLKGETRHINELLKDWFAQIGLNLIDFKLEFG 180  
 YKND+L+DPPFINDEHVKFLGIVNDEEIAYLKGETR INELLK WFAQIGLNLIDFKLEFG  
 Sbjet: 131 YKNDLDDPFINDEHVKFLGIVNDEEIAYLKGETRRINELLKGWFAQIGLNLIDFKLEFG 190

40 Query: 181 FDKDGKIILADEFSPDNCRLWDADGNHMDKDVFRDLGSLTDVYQVVLEKLIAL 234  
 FD++G IILADEFSPDNCRLWD +GNHMDKDVFRDLG+LTDVYQVVLEKLIAL  
 Sbjet: 191 FDQEGTIILADEFSPDNCRLWDKNGNHMDKDVFRDLGNLTDVYQVVLEKLIAL 244

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 vaccines or diagnostics.

#### 45 Example 306

A DNA sequence (GBSx0335) was identified in *S.agalactiae* <SEQ ID 987> which encodes the amino acid  
 sequence <SEQ ID 988>. Analysis of this protein sequence reveals the following:

Possible site: 13  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.2779(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9457> which encodes amino acid sequence <SEQ ID 9458>  
 was also identified.

The protein has homology with the following sequences in the GENPEPT database:



-391-

>GP:AAC35700 GB:AF041468 acyl carrier protein [Guillardia theta]  
Identities = 27/75 (36%), Positives = 52/75 (69%)

5 Query: 12 MSRDEVFEKMLELLRQQLGDPQLDITPESSLHDDLAI DSIALTEFIINLEDV FHLEIPDE 71  
M+ E+FEK+ ++ +QLG + +T +++ +DL DS+ E ++ +E+ F++EIPD+  
Sbjct: 1 MNEQEIFEKVQTIISEQLGVDSQVTKDANFANDLGADSLDTVELVMAIEEAFNIEIPDD 60

Query: 72 AVEHMSSVQQLLDYI 86  
A E +S++QQ +D+I  
10 Sbjct: 61 AAEQISNLQQAVDFI 75

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 989> which encodes the amino acid sequence <SEQ ID 990>. Analysis of this protein sequence reveals the following:

15 Possible site: 24  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.1917(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 36/77 (46%), Positives = 57/77 (73%)

25 Query: 12 MSRDEVFEKMLELLRQQLGDPQLDITPESSLHDDLAI DSIALTEFIINLEDV FHLEIPDE 71  
M+R E+FE+++ L+++Q + IT ++ L +DLA+DSI L EFIIN+ED FH+ IPDE  
Sbjct: 1 MTRQEIFERLINLIQKQRSYLSVAITEQTHLKNDLAVDSIELVEFIINVEDEFHIAIPDE 60

Query: 72 AVEHMSSVQQLLDYIIE 88  
VE M ++ +LDY+++  
30 Sbjct: 61 DVEDMVFMRDILDYLVQ 77

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 35 Example 307

A DNA sequence (GBSx0336) was identified in *S.agalactiae* <SEQ ID 991> which encodes the amino acid sequence <SEQ ID 992>. This protein is predicted to be fatty acid/phospholipid synthesis protein (plsX). Analysis of this protein sequence reveals the following:

40 Possible site: 21  
>>> Seems to have no N-terminal signal sequence  
INTEGRAL Likelihood = -0.64 Transmembrane 101 - 117 ( 101 - 117)

----- Final Results -----  
bacterial membrane --- Certainty=0.1256(Affirmative) < succ>  
45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9455> which encodes amino acid sequence <SEQ ID 9456> was also identified.

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB13462 GB:Z99112 alternate gene name: ylpD [Bacillus subtilis]  
Identities = 174/329 (52%), Positives = 238/329 (71%), Gaps = 2/329 (0%)

55 Query: 8 KIAIDAMGGDYAPKAIVEGVNQAISDFSDIEVQLYGDQKKIEKYLTVT-ERVSIHTEEK 66  
+IA+DAMGGD+APKA+++GV + I F D+ + L GD+ IE +LT T +R++++H +E  
Sbjct: 2 RIAVDAMGGDHAPKAVIDGVIKGIEAFDDLHITLVGDKTTIESHLTTTSDRITVLHADEV 61

-392-

Query: 67 INSDDEPAKAVRRKKQSSMVLGAKAVKDGVAQAFISAGNTGALLAAGLFVVGRIKGVDRP 126  
 I DEP +AVRRKK SSMVL A+ V + A A ISAGNTGAL+ AGLF+VGRIKG+DRP  
 Sbjct: 62 IEPTDEFPVRAVRRKKNSSMVLMAQEVANRADACISAGNTGALMTAGLFTVGRIKIDRP 121

Query: 127 GLMSTMPITLDGVGFMDLDLGANAENTASHLHQYAILGSFYAKNVRGIEVPRVGLLNNGTE 186  
 L T+PT+ G GF +LD+GAN + HL QYAI+GS Y++ VRG+ PRVGLLN GTE  
 Sbjct: 122 ALAPTLPVTSVGDGFLLLDVGANVDAKPEHLVQYAIMGSVYSQQVRGVTSPRVGLLNNGTE 181

Query: 187 ETKGDSLHKEAYELLAAEPSINFIGNIEARDLMSSVADVVDGFTGNAVLTMEGTAMS 246  
 + KG+ L K+ +++L +INFIGN+EARDL+ VADVVDGFTGN LKT+EG+A+S  
 Sbjct: 182 DKKGNELTKQTFQILKETANINFIGNVEARDLLDDVADVVDGFTGNVTLKTLGSALS 241

Query: 247 IMGSLKSSIKSGGVKAKLGALLLKDSLYQLKDSMDYSSAGGAVLFGKAPIVKCHGSSDS 306  
 I ++ + + +KL A +LK L ++K M+YS+ GGA LFGLKAP++K HGSSDS  
 Sbjct: 242 IFKMMR-DVMTSTLTSKLAAAVLKPKLKEMKMKMEYSNYGGASLFGLKAPVIKAHGSSDS 300

Query: 307 KAVYSTLKQVRTMLETQVVDQLVDAFTDE 335  
 AV+ ++Q R M+ V + + +E  
 Sbjct: 301 NAVFHAIRQAREMVSQNVAALIQEEVKEE 329

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 993> which encodes the amino acid sequence <SEQ ID 994>. Analysis of this protein sequence reveals the following:

Possible site: 36  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -2.07 Transmembrane 121 - 137 ( 120 - 138)

----- Final Results -----  
 bacterial membrane --- Certainty=0.1829(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9127> which encodes the amino acid sequence <SEQ ID 9128>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 16  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -2.07 Transmembrane 95 - 111 ( 94 - 112)

----- Final Results -----  
 bacterial membrane --- Certainty= 0.183(Affirmative) < succ>  
 bacterial outside --- Certainty= 0.000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 254/330 (76%), Positives = 290/330 (86%)

Query: 6 MKKIAIDAMGGDYAPKAIVEGVNQAI SDFSIEVQLYGDQKKIEKYLTVTERVSIHTTE 65  
 MK+IAIDAMGGD APKAIVEGVNQAI FSDIE+QLYGDQ KI YL ++RV+IIHT+E  
 Sbjct: 27 MKRIAIDAMGGDNAPKAIVEGVNQAI EAFSDIEIQLYGDQTKINSYLIQSDRVAIHTDE 86

Query: 66 KINSDDDEPAKAVRRKKQSSMVLGAKAVKDGVAQAFISAGNTGALLAAGLFVVGRIKGVDR 125  
 KI SDDEPAKAVRRKK++SMVL AKAVK+G A A ISAGNTGALLA GLFVVGRIKGVDR  
 Sbjct: 87 KIMSDDDEPAKAVRRKKKASMVLA AKAVKEGKADAIISAGNTGALLAVGLFVVGRIKGVDR 146

Query: 126 PGLMSTMPITLDGVGFMDLDLGANAENTASHLHQYAILGSFYAKNVRGIEVPRVGLLNNGT 185  
 PGL+ST+PT+ G+GFDMLDLGANAENTA HLHQYAILGSFYAKNVRGI PRVGLLNNGT  
 Sbjct: 147 PGLLSTIPTVTGLGFDMLDLGANAENTAKHLHQYAILGSFYAKNVRGIANPRVGLLNNGT 206

Query: 186 EETKGDSLHKEAYELLAAEPSINFIGNIEARDLMSSVADVVDGFTGNAVLTMEGTAM 245  
 EETKGD L K YELL A+ +I+F+GN+EAR+LMS VADV+V+DGFTGNAVLK++EGTA+  
 Sbjct: 207 EETKGDPLRKATYELLTADNTISFVGNVEARELMMSGVADVIVSDGFTGNAVLSIEGTAI 266

Query: 246 SIMGSLKSSIKSGGVKAKLGALLLKDSLYQLKDSMDYSSAGGAVLFGKAPIVKCHGSSD 305

-393-

SIMG LK I SGG+K K+GA LLK SLY++K ++DYSSAGGAVLFGLKAP+VK HGSSD  
 Sbjct: 267 SIMGQLKQIINSGGIKTKIGASLLKSSLYEMKKTLDYSSAGGAVLFGLKAPVVKSHGSSD 326

Query: 306 SKAVYSTLKQVRTMLETQVVDQLVDAFTDE 335  
 KA++ST+KQVRTML+T VV QLV+ F E  
 Sbjct: 327 VKAIFSTIKQVRTMLDTNVVGQLVEEFAKE 356

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 308

A DNA sequence (GBSx0337) was identified in *S.agalactiae* <SEQ ID 995> which encodes the amino acid sequence <SEQ ID 996>. Analysis of this protein sequence reveals the following:

Possible site: 27  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.4668(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 309

A DNA sequence (GBSx0338) was identified in *S.agalactiae* <SEQ ID 997> which encodes the amino acid sequence <SEQ ID 998>. Analysis of this protein sequence reveals the following:

Possible site: 32  
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -12.84	Transmembrane	61 - 77 ( 55 - 82)
INTEGRAL	Likelihood = -10.14	Transmembrane	26 - 42 ( 19 - 51)
INTEGRAL	Likelihood = -9.77	Transmembrane	192 - 208 ( 186 - 211)
INTEGRAL	Likelihood = -5.79	Transmembrane	267 - 283 ( 262 - 286)
INTEGRAL	Likelihood = -3.77	Transmembrane	100 - 116 ( 99 - 116)

----- Final Results -----  
           bacterial membrane --- Certainty=0.6137(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9453> which encodes amino acid sequence <SEQ ID 9454> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA22372 GB:AL034446 putative transmembrane protein  
 [Streptomyces coelicolor A3(2)]  
 Identities = 47/154 (30%), Positives = 69/154 (44%), Gaps = 12/154 (7%)

Query: 120 SGFVEISSNSFSFGPFFFLFLAYFIQSLTEEILFRGYVMTTVTKFKGSFAGVLCNSMLF 179  
 SG+ E+ S F+A + TEE++FRG + + + G++ + ++F  
 Sbjct: 118 SGYYEVDGLGSVQGAIGLVGFMA--AAATEEVVFRGVLFRIIEEHIGTYLALGLTGLVF 175

Query: 180 SFIHFRN-----YGITAIALFNLFLGLIIFSILEFNMTKNILFVTGVHTTNFTMGCVLGN 234  
 +H N +G AIA+ F+L ++ T+N+ GVH WNF G V

-394-

Sbjct: 176 GLMHLLNEDATLWGALAIATEAGFMLAAAYAA----TRNLWLTIGVHFGWNFAAGGVFST 231

Query: 235 KVSGGDSVPVSLFRITENSSFALWNGGDFGFEGGV 268

VSG L T S L GGDFG EG V

Sbjct: 232 VVSGNGDSEGLLDAT-MSGPKLLTGDFGPEGSV 264

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 310

A DNA sequence (GBSx0339) was identified in *S.agalactiae* <SEQ ID 999> which encodes the amino acid sequence <SEQ ID 1000>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2665(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9451> which encodes amino acid sequence <SEQ ID 9452> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB05088 GB:AP001511 unknown conserved protein [Bacillus halodurans]  
Identities = 81/242 (33%), Positives = 124/242 (50%), Gaps = 3/242 (1%)

Query: 8 GLVLYNRNYREDDKLVKIFTETEGKRMFFVKHAS--KSKFNAVLQPLTIAHFILKINDNG 65  
G+V+ +Y E +K+V +FT GK + A KS+ AV Q T + + N G

Sbjct: 7 GIVIRTVDYGESNKIVTVFTREYGKIALMARGAKRPKSRLTAVTQLFTYGMFMFQKNA-G 65

Query: 66 LSYIDDYKEVLAFQETNSDLFKLSYASYITSLADVAISDNVADAQLFIFLKKTLIELIEDG 125  
L + + + +F+E +DLF+ SY SY+T L + D + LF L +T+ + +G

Sbjct: 66 LGTLTQGEIIQSFREVRNDLFRASYVSVYTDLTNKLTEDEKRNPLYFELLYQTIHYMNEG 125

Query: 126 LDYEILTNIFEVQLLERFGVALNFHDCVFCRVGLPFDPSHKYSGLLCPNHYYKDERRNH 185  
+D ++LT IFEV++ G+ CV C +P FS K +G LC KD

Sbjct: 126 MDPDVLTRIFEVKMFTVAGIKPELDQCVCSCSTDVPVGFSTKEAGFLCKRCIEKDPHAYK 185

Query: 186 LDPNMLYLINRFQSIQFDDLTISVKPEMKLKIRQFLDMIYDEYVGIHLKSKKFIDDLSSWG 247  
+ + L+ F L TIS+KPE K ++ + YDEY G+HLKS++F+D L S G

Sbjct: 186 ITAQVAKLLRLFYHFDLQRLGTISLKPETKATLKTIIHQYYDEYSGLHLKSRFLDQLESMD 247

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1001> which encodes the amino acid sequence <SEQ ID 1002>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1566(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 159/251 (63%), Positives = 210/251 (83%)

Query: 1 MRVSQTYGLVLYNRNYREDDKLVKIFTETEGKRMFFVKHASKSKFNAVLQPLTIAHFILK 60

-395-

M++++ G+VL+NRNYREDDKLVKIFTE GK+MFFVKH S+SK ++++QPLTIA FI K  
 Sbjct: 1 MQLTESLGIVLFNRNYREDDKLVKIFTEVAGKQMFVKHISRSMSSIIQPLTIADFIFK 60

Query: 61 INDNGLSYIDYKEVLAFQETNSDLFKLSYASYITSLADVAISDNVADAQLFIFLKKTTLE 120  
 +ND GLSY+ DY V ++ N+D+F+L+YASY+ +LAD AI+DN +D+ LF FLKKTL+  
 Sbjct: 61 LNDTGLSYVVDYSNVNTYRYINNDIFRLAYASYVLALADAAIADNESDSHLFTFLKKTLD 120

Query: 121 LIEDGLDYEILTNI FEVQLLERFGVALNFHDCVFCRVRGLPFD FSHKYSGLLCPNHYYKD 180  
 L+E+GLDYEILTNI FE+Q+L+RFG++LNFH+C CHR LP DFSH++S +LC HYYKD  
 Sbjct: 121 LMEEGLDYEILTNI FEIQILDRFGISLNFHECAICHR TDLP LDFSHRFSAVLCSEHYYKD 180

Query: 181 ERRNHLDPNMLYLINRFQSIQFDDLTISVKPEMKLKIRQFLDMIYDEYVGIHLKSKKFI 240  
 RRNHLDPN++YL++RFQ I FDDL+TIS+ ++K K+RQF+D +Y +YVGI LKSK FI  
 Sbjct: 181 NRRNHLDPNVIYLLSRFQKITFDDLRTISLNKDIKKLRQFIDELYHDYVGIKLSKSTFI 240

Query: 241 DDLSSWGSIMK 251  
 D+L WG IMK  
 Sbjct: 241 DNLVKWGDIMK 251

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 311

A DNA sequence (GBSx0340) was identified in *S.agalactiae* <SEQ ID 1003> which encodes the amino acid sequence <SEQ ID 1004>. This protein is predicted to be aromatic amino acid aminotransferase (patA). Analysis of this protein sequence reveals the following:

Possible site: 14  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -3.13 Transmembrane 141 - 157 ( 140 - 159)

30 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2253(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35 A related GBS nucleic acid sequence <SEQ ID 9449> which encodes amino acid sequence <SEQ ID 9450> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF06954 GB:AF146529 aromatic amino acid aminotransferase  
 [Lactococcus lactis subsp. cremoris]  
 40 Identities = 261/391 (66%), Positives = 323/391 (81%)

Query: 38 MTLEKRFNKYLDRIEVSILIRQFDQSISDIPGMVKLTLGEPDFTTPDHVKEAKSAIDANQ 97  
 M L K+FN LD+IE+SLIRQFDQ +S IP ++KLTLGEPDF TP+HVK+A +AI+ NQ  
 Sbjct: 1 MDLLKKFNPNLDKIEISLIRQFDQVSSIPDIKLTGEPDFYTPHEVVKQAGIAAIENNQ 60

45 Query: 98 SYTGMGSLLLRQAAADFADKYNLTYPDCEILVTIGATEALSASLIAILEAGDVVLL 157  
 S+YTGM+GLL LRQAA++F KY L+Y + EILVT+G TEA+S+ L++IL AGD VL+  
 Sbjct: 61 SHYTGMAGLLELRQAAEFLLKKYGLSYAAEDEILVTGVTEAISSVLLSILVAGDEVLI 120

50 Query: 158 PAPAYPGYEPVNLVGADIVEIDTRENDRLTPPEMLETAIIQQGEKLVKAVLLNYPNPTG 217  
 PAPAYPGYEP++ L G +VEIDTR NDF LTPPEML+ AII++ K+KAV+LNYP NPTG  
 Sbjct: 121 PAPAYPGYEPLITLAGGSLVEIDTRANDFVLTPEMLDQAIIEREGVKAVILNYPANPTG 180

55 Query: 218 ITYSRQEI AALAEVLKKYDIFVISDEVYSELTYTGQHVSI AEYLPNQ TILNGLSKSHA 277  
 +TY+R++I LAEVLKK+++FVI+DEVYSEL YT Q HVSTAEY P QTI++NGLSKSHA  
 Sbjct: 181 VTYNREQIKDLAEVLKKHEVFVIADEVYSELNYTDQPHVSTAEYAPEQTIVL NGLSKSHA 240

Query: 278 MTGWRVGLVYAPEAFIAQIIKSHQYMTAASTISQFAGVEALSVGKNDTLP MRQGYIKRR 337  
 MTGWR+GL++A +AQIIK+HQY+VT+AST SQFA +EAL G +D LPM++ Y+KRR

-396-

Sbjct: 241 MTGWRIGLIFAARELV AQIIKTHQYLVTSASTQSQFAATEALKNGADDALPMKKEYLKRR 300

Query: 338 DYIIDKMSKLGFKIIPSGAFYIFAKIPDSYPQDSFKFCQDFAYQQAVAIIPGVAFGKYG 397  
DYII+KMS LGFKII+P GAFYIFAKIP QDSFKF DFA + AVAIIPG+AFG+YG

5 Sbjct: 301 DYIIEKMSALGFKIIEPDGAFYIFAKIPADLEQDSFKFAVDFAKENAVAIIPGIAFGQYG 360

Query: 398 EGYIRLSYAASMEVIETAMARLKVFMESEYEG 428  
EG++RLSYAASM+VIE AMARL ++ G

10 Sbjct: 361 EGFVRLSYAASMDVIEQAMARLTDYVTKKRG 391

There is also homology to SEQ ID 1006.

SEQ ID 1004 (GBS332) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 3; MW 50.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 67 (lane 4; MW 76kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 312

A DNA sequence (GBSx0341) was identified in *S.agalactiae* <SEQ ID 1007> which encodes the amino acid sequence <SEQ ID 1008>. This protein is predicted to be ribose-phosphate pyrophosphokinase (prsA).

Analysis of this protein sequence reveals the following:

Possible site: 22  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3118(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9447> which encodes amino acid sequence <SEQ ID 9448> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA62181 GB:M92842 prs [Listeria monocytogenes]  
Identities = 209/312 (66%), Positives = 266/312 (84%), Gaps = 3/312 (0%)

35 Query: 10 LKLFALSSNKELAKKVSQTIGIPLGQSTVRQFSDGEIQVNIEESIRGHHVFILQSTSSPV 69  
LK+F+L+SN+ELA++++ +GI LG+S+V FSDGEIQ+NIEESIRG HV+++QSTS+PV  
Sbjct: 10 LKIFSLNSNRELAEEIAKEVGIELGKSSVTHFSDGEIQINIEESIRGCHVYVIQSTSNPV 69

40 Query: 70 NDNLMELIMVDALKRASAESVSVMPPYGYARQDRKARSREPITSKLVANMLEVAGVDR 129  
N NLME+LIM+DALKRASA ++++VMPYGYARQDRKARSREPIT+KLVAN++E AG R  
Sbjct: 70 NQNLMEELIMIDALKRASAATINIVMPYGYARQDRKARSREPITAKLVANLIETAGATR 129

45 Query: 130 LLTVDLHAAQIQGFFDIPVDHLMGAPLIADYFDRQGLVGDDVVVSPDHGGVTRARKLAQ 189  
++T+D+HA QIQGFFDIP+DHL L++DYF + L GDD+VVVSPDHGGVTRARK+A  
Sbjct: 130 MITLDMHAPQIQGFFDIPIDHLNAVRLSDYFSEHRL-GDDLVVVSPDHGGVTRARKMAD 188

50 Query: 190 CLKTPIAIIDKRRSVTKMNTSEVMNIIGNIKGKCCILIDDMIDTAGTICHAADALAEAGA 249  
LK PIAIIDKRR + N +EVMNI+GN++GK CI+IDD+IDTAGTI AA AL EAGA  
Sbjct: 189 RLKAPIAIIIDKRR--PRPNVAEVMNIVGNVEGKVCIIIDDIIDTAGTITLAAKALREAGA 246

55 Query: 250 TAVYASCTHPVLSPALDNIQNSAIEKLIVLDTIYLPEERLIDKIEQISTAEELIGEAIIR 309  
T VYA C+HPVLSPGA+ I+ S IEKL+V ++I LP EE+ IDK+EQ+S+A L+GEAI+R  
Sbjct: 247 TKVYACCSHPVLSPGAMKRIEESPIEKLIVVTNSIALPEEKWIDKMEQLSVAALLGEAIVR 306

Query: 310 IHEKRPLSPLFE 321  
+HE +S LFE

-397-

Sbjct: 307 VHNASVSSLFE 318

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1009> which encodes the amino acid sequence <SEQ ID 1010>. Analysis of this protein sequence reveals the following:

5 Possible site: 22  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2685(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 298/322 (92%), Positives = 311/322 (96%)

15 Query: 1 MEEIMSYSNLKL FALSSNKELAKV SQTIGIPLGQSTVRQFSDGEIQVNIEESIRGHHVF 60  
+EE MSYS+LKL FALSSNKELA+KV+ +GI LG+STVRQFSDGEIQVNIEESIRGHHVF  
Sbjct: 1 LEEKMSYSDLKL FALSSNKELAEKVASAMGIQLGKSTVRQFSDGEIQVNIEESIRGHHVF 60

20 Query: 61 ILQSTSSPVNDNLMEILIMVDALKRASAE SVSVMPYYGYARQDRKARSREPITSKLVAN 120  
ILQSTSSPVNDNLMEILIMVDALKRASAE +SVMPYYGYARQDRKARSREPITSKLVAN  
Sbjct: 61 ILQSTSSPVNDNLMEILIMVDALKRASAEKISVMPYYGYARQDRKARSREPITSKLVAN 120

25 Query: 121 MLEVAGVDRLLTVDLHAAQIQGFFDIPVDHLMGAPLIADYFDRQGLVGDDVVVVSPDHGG 180  
MLEVAGVDRLLTVDLHAAQIQGFFDIPVDHLMGAPLIADYFDR GLVG+DVVVVSPDHGG  
Sbjct: 121 MLEVAGVDRLLTVDLHAAQIQGFFDIPVDHLMGAPLIADYFDRHGLVGEDVVVVSPDHGG 180

30 Query: 181 VTRARKLAQCLKTPIAIIIDKRRSVTKMNTSEVMNIIGNIKGKKCILIDDMIDTAGTICHA 240  
VTRARKLAQ L+TPIAIIIDKRRSV KMNTSEVMNIIGN+ GKKCILIDDMIDTAGTICHA  
Sbjct: 181 VTRARKLAQFLQTPIAIIIDKRRSVDKMNTSEVMNIIGNVSGKKCILIDDMIDTAGTICHA 240

35 Query: 241 ADALAEAGATAVYASCTHPVLSGPALDNIQNSAIEKLIVLDTIYLPPEERLIDKIEQISIA 300  
ADALAEAGATAVYASCTHPVLSGPALDNIQ SAIEKLIVLDTIYLP+ERLIDKIEQISIA  
Sbjct: 241 ADALAEAGATAVYASCTHPVLSGPALDNIQSAIEKLIVLDTIYLPKERLIDKIEQISIA 300

Query: 301 ELIGEAIIRIHEKRPLSPLFEM 322  
+L+ EAIIRIHEKRPLSPLFEM  
Sbjct: 301 DLVAEAIIRIHEKRPLSPLFEM 322

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 313

45 A DNA sequence (GBSx0342) was identified in *S.galactiae* <SEQ ID 1011> which encodes the amino acid sequence <SEQ ID 1012>. This protein is predicted to be a secreted protein. Analysis of this protein sequence reveals the following:

Possible site: 20  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.3751(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 A related GBS nucleic acid sequence <SEQ ID 9277> which encodes amino acid sequence <SEQ ID 9278> was also identified.

The protein has homology with the following sequences in the GENPEPT database:





[illegible]

SEQ ID 8544 (GBS65) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 5 (lane 6; MW 47.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 3; MW 72kDa) and in Figure 175 (lane 2 & 3; MW 72kDa).

The GBS65-GST fusion product was purified (Figure 102A; see also Figure 191, lane 4) and used to immunise mice (lane 1 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 102B), FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 314

A DNA sequence (GBSx0343) was identified in *S.agalactiae* <SEQ ID 1015> which encodes the amino acid sequence <SEQ ID 1016>. Analysis of this protein sequence reveals the following:

Possible site: 18

```
>>> Seems to have no N-terminal signal sequence
```

----- Final Results' -----

bacterial cytoplasm --- Certainty=0.1184 (Affirmative) < succ>

'bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-400-

**Example 315**

A DNA sequence (GBSx0344) was identified in *S.agalactiae* <SEQ ID 1017> which encodes the amino acid sequence <SEQ ID 1018>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4736(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 316**

A DNA sequence (GBSx0345) was identified in *S.agalactiae* <SEQ ID 1019> which encodes the amino acid sequence <SEQ ID 1020>. This protein is predicted to be elongation factor Tu (tufA). Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3012(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9737> which encodes amino acid sequence <SEQ ID 9738> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB03851 GB:AP001507 translation elongation factor Tu (EF-Tu)

[Bacillus halodurans]

Identities = 302/397 (76%), Positives = 350/397 (88%), Gaps = 2/397 (0%)

Query: 7 MAKEKYDRSKPHVNIQTIGHVDHGKTTTLTAAITTVLARRLPTSVNQPKDYASIDAAPPEER 66

MAKEK+DRSK H NIGTIGHVDHGKTTTLTAAITTVLA+R V Y +ID APEER

Sbjct: 1 MAKEKFDRSKTHANIGTIGHVDHGKTTTLTAAITTVLAKRSGKGVAMA--YDAIDGAPEER 58

Query: 67 ERGITINTAHVEYETEKRHYAHIDAPGHADYVKNMITGAAQMDGAILVVASTDGPMPQTR 126

ERGITI+TAHVEYET+ RHYAH+D PGHADYVKNMITGAAQMDG ILVV++ DGPMPQTR

Sbjct: 59 ERGITISTAHVEYETDNRYAHVDCPGHADYVKNMITGAAQMDGGILVVSADGPMQTR 118

Query: 127 EHILLSRQVGKHLIVFMNKKVDLVDDEELLELVEMEIRDLLSEYDFPGDDLPIVQGSALK 186

EHILLSRQVGK +L+VF+NK D+VDDEELLELVEME+RDLLSEYDFPGDD+PVI+GSALK

Sbjct: 119 EHILLSRQVGVPYLVVFLNKCMDVDEELLELVEMEVRDLLSEYDFPGDDVPVIRGSALK 178

Query: 187 ALEGDEKYEDIIMELMSTVDEYIPEPERDTPKPLLLPVEDVFSITGRGTVAASGRIDRGTV 246

ALEGD ++E+ I+ELM+ VD+YIP PERDT+KP ++PVEDVFSITGRGTVA+GR++RG +

Sbjct: 179 ALEGDAEWEEKIIEIEMAVIDDIPTPERDTEKPFMPVEDVFSITGRGTVAATGRVERGQL 238

Query: 247 RVNDEVEIVGIKEDIQKAVVTGVEMFRKQLDEGLAGDNVGVLLRGVQRDEIERGQVLAKP 306

V DEVEI+G++E+ +K VTGVEMFRK LD AGDN+G LLRGV R+E++RGQVLAKP

Sbjct: 239 NVGDEVEIIGLEEEAKKTTVTGVEMFRKLLDYAEAGDNIGALLRGVSREEVQRGQVLAKP 298

-401-

Query: 307 GSINPHTRFKGEVYILSKEEGGRHTPPFFNNYRPQFYFRITDVTGSIELPAGTEMVMPGDN 366  
 G+I PHT FK EVY+LSKEEGGRHTPPFF+NYRPQFYFRITDVTG I+LP G EMVMPGDN  
 Sbjct: 299 GTITPHTNFKAENVYLSKEEGGRHTPPFFSNYRPQFYFRITDVTGIIQLPDGVEMVMPGDN 358

5 Query: 367 VTIEVELIHPIAVEQGTTFSSIREGGRTVGSGIVSEIE 403  
 V + VELI PIA+E+GT FSIREGGRTVG+G+V+ I+  
 Sbjct: 359 VEMTVELIPIAIEEGTKFSIREGGRTVGAGVVASIQ 395

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1021> which encodes the amino acid  
 10 sequence <SEQ ID 1022>. Analysis of this protein sequence reveals the following:

Possible site: 19  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.1367(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

20 Identities = 386/404 (95%), Positives = 396/404 (97%)

Query: 1 MEAFPKMAKEKYDRSKPHVNIGTIGHVDHGKTTTLTAAITTVLARRLPTSVNQPKDYASID 60  
 +EAFPKMAKEKYDRSKPHVNIGTIGHVDHGKTTTLTAAITTVLARRLP+SVNQPKDYASID  
 Sbjct: 12 LEAFPKMAKEKYDRSKPHVNIGTIGHVDHGKTTTLTAAITTVLARRLPSSVNQPKDYASID 71

25 Query: 61 AAPEERERGITINTAHVEYETEKRHYAHIDAPGHADYVKNMITGAAQMDGAILVVASTDG 120  
 AAPEERERGITINTAHVEYET RHYAHIDAPGHADYVKNMITGAAQMDGAILVVASTDG  
 Sbjct: 72 AAPEERERGITINTAHVEYETATRHYAHIDAPGHADYVKNMITGAAQMDGAILVVASTDG 131

30 Query: 121 PMPQTRHILLSRQGVGKHLIVFMNKVDLVDDEELLELVEMEIRDLLSEYDFPGDDLPIV 180  
 PMPQTRHILLSRQGVGKHLIVFMNKVDLVDDEELLELVEMEIRDLLSEYDFPGDDLPIV  
 Sbjct: 132 PMPQTRHILLSRQGVGKHLIVFMNKVDLVDDEELLELVEMEIRDLLSEYDFPGDDLPIV 191

35 Query: 181 QGSALKALEGDEKEYEDIIMELMSTVDEYIPEPERDTDKPLLLPVEDVFSITGRGTVASGR 240  
 QGSALKALEGD K+EDIIMELM TVD YIPEPERDTDKPLLLPVEDVFSITGRGTVASGR  
 Sbjct: 192 QGSALKALEGDTKFEDIIMELMDTVDSYIPEPERDTDKPLLLPVEDVFSITGRGTVASGR 251

40 Query: 241 IDRGTVRVNDEVEIVGIKEDIQKAVVTGVEMFRKQLDEGLAGDNVGVLLRGVQRDEIERG 300  
 IDRGTVRVNDE+EIIVGIKE+ +KAVVTGVEMFRKQLDEGLAGDNV+LLRGVQRDEIERG  
 Sbjct: 252 IDRGTVRVNDEIEIVGIKEETKAVVTGVEMFRKQLDEGLAGDNVGIILLRGVQRDEIERG 311

45 Query: 301 QVLAKPGSINPHTRFKGEVYILSKEEGGRHTPPFFNNYRPQFYFRITDVTGSIELPAGTEM 360  
 QV+AKP SINPHT+FKGEVYILSK+EGGRHTPPFFNNYRPQFYFRITDVTGSIELPAGTEM  
 Sbjct: 312 QVIAPSSINPHTKFKGEVYILSKDEGGRTTPFFNNYRPQFYFRITDVTGSIELPAGTEM 371

Query: 361 VMPGDNVTIEVELIHPIAVEQGTTFSSIREGGRTVGSGIVSEIEA 404  
 VMPGDNVTI VELIHPIAVEQGTTFSSIREGGRTVGSGIVSEIEA  
 Sbjct: 372 VMPGDNVTINVELIHPIAVEQGTTFSSIREGGRTVGSGIVSEIEA 415

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 vaccines or diagnostics.

### Example 317

A DNA sequence (GBSx0346) was identified in *S.agalactiae* <SEQ ID 1023> which encodes the amino  
 acid sequence <SEQ ID 1024>. Analysis of this protein sequence reveals the following:

55 Possible site: 36  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -0.64 Transmembrane 90 - 106 ( 90 - 106)

----- Final Results -----

60 bacterial membrane --- Certainty=0.1256(Affirmative) < succ>

-402-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 318

10 A DNA sequence (GBSx0347) was identified in *S.agalactiae* <SEQ ID 1025> which encodes the amino acid sequence <SEQ ID 1026>. This protein is predicted to be ftsW. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have no N-terminal signal sequence

15 INTEGRAL Likelihood = -11.15 Transmembrane 44 - 60 ( 35 - 70)  
 INTEGRAL Likelihood = -4.73 Transmembrane 76 - 92 ( 74 - 98)  
 INTEGRAL Likelihood = -3.88 Transmembrane 117 - 133 ( 113 - 134)

----- Final Results -----

20 bacterial membrane --- Certainty=0.5458 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

25 >GP:AAB39929 GB:U58049 putative cell division protein ftsW  
 [Enterococcus hirae]  
 Identities = 78/159 (49%), Positives = 107/159 (67%), Gaps = 4/159 (2%)  
 Query: 1 MANSXYAMSGGWFGRLGNSIEKLGYPLEATTDFVFSIVIEELGVIGAGFILALVFFLI 60  
 M+NS YA+ NGG FGRG+GNSI K GYLPE+ TDF+FS++ EE G+IGA +L L+F L  
 30 Sbjct: 240 MSNSYYALYNGGLFGRGMGNSITKKGYLPESETDFIFSVIAEEFLIGALLVLFLLFLLC 299  
 Query: 61 LRIMHVGIKADPFNSMIALGIGAMLLMQVFVNIGGISGLIPSTGVTFPFLSQGGNSLLV 120  
 +RI K K+ ++I +G+G +L+Q +NIG I GLIP TGV PF+S GG S L+  
 35 Sbjct: 300 MRIFQKSTKQKNQANLILIGVGTWILVQTSINIGSILGLIPMTGVPLPFVSYGGTSYLI 359  
 Query: 121 LSVAIGFVLNIDANEKKELIMKEAEEQYKPKQEKNEKIIN 159  
 LS AIG LNI + + KE + + + Q K K++N  
 Sbjct: 360 LSFAIGLALNISSRQVKE----KNKQVERLQLKKPKLLN 394

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1027> which encodes the amino acid sequence <SEQ ID 1028>. Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

45 INTEGRAL Likelihood = -10.93 Transmembrane 312 - 328 ( 303 - 338)  
 INTEGRAL Likelihood = -8.23 Transmembrane 22 - 38 ( 17 - 47)  
 INTEGRAL Likelihood = -6.85 Transmembrane 192 - 208 ( 187 - 211)  
 INTEGRAL Likelihood = -5.10 Transmembrane 218 - 234 ( 212 - 236)  
 INTEGRAL Likelihood = -4.83 Transmembrane 86 - 102 ( 85 - 107)  
 50 INTEGRAL Likelihood = -3.72 Transmembrane 385 - 401 ( 383 - 402)  
 INTEGRAL Likelihood = -3.45 Transmembrane 61 - 77 ( 61 - 79)  
 INTEGRAL Likelihood = -2.39 Transmembrane 344 - 360 ( 344 - 360)

----- Final Results -----

55 bacterial membrane --- Certainty=0.5373 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB59721 GB:AJ250603 FtsW protein [Enterococcus faecium]

Identities = 131/397 (32%), Positives = 223/397 (55%), Gaps = 23/397 (5%)

5 Query: 15 KRHLNYSILLPYLILSVIGLIMVYSTTSVSLIQAHANPFKSVINQGVFWIISLVAITFI 74  
KR +++ IL PYL LS+IGL+ VYS +S L+QA N ++ Q +F +S I  
Sbjct: 3 KRKKIDWWILGPYLTLISMIGLLEVYSASSYRLQADENTKSLLLRQLIFIFLSWGVIFLA 62

10 Query: 75 YKLKLNFLINTRVLTVMVLGEAFLLIAR--FFTTAIKGAHGWIIVIGPVSFQPAEYLKII 132  
+KL++L + ++ + F LI+ R F + GA WI + + FQP+E +  
Sbjct: 63 RSIKHLHYLLHPKIAGYGLALSIFFLILVRVGIFGVTVNGAQRWISLFGIQFQPSELANLF 122

15 Query: 133 MVWYLALTFAKIQKNISLYDYQALTRRKWWPTQWNDLRDWRVYSLMLVLLVAAQPDIGNA 192  
+++YL+ F P + +L+ + ++ + LL+ QP + A  
Sbjct: 123 LIFYLSWFFRDGNN-----PPK--NLKKPFLITVSITLLILFQPKIAGA 164

20 Query: 193 SIIVLTAIMFSISIGIGYRWFSAILVMITGLSTVFLGTIAVIGVERVAKIP-VFGYVAKR 251  
+I+ A ++F + + ++ ++V + L G + +G + +P +F + +R  
Sbjct: 165 LMILSIWVIFWAAAVPFKKGIYLVTFSAALLIGAAGGVLYLGNK--GWLPMFNHAYER 222

25 Query: 252 FSAFFNPFHDLTDSGHQLANSYYAMSNGGWFGQGLGNSIEKRGYLPQAQTDVFVSVVIEE 311  
+ +PF D +G+Q+ +S+YA+ NGG +G+GLGNSI K+GYLPE +TDF+FS++ EE  
Sbjct: 223 IATLRDPFIDSHGAGYQMTSHFYALYNGGIWGRGLGNSITKKGYPETETDFIFSIITEE 282

30 Query: 372 TGVTFPFLSQGNSLLVLSVAVGFVLNIDASEKRDDI 408  
TGV PF+S GG S L+LS+ +G LNI + + +++  
Sbjct: 343 TGVPLPFVSYGGTSYLILSLGIGITLNISSKIQAEL 379

An alignment of the GAS and GBS proteins is shown below:

Identities = 130/166 (78%), Positives = 152/166 (91%), Gaps = 2/166 (1%)

35 Query: 1 MANSXYAMSNGGWFGRLGNSIEKLGYLPEATTDFVFSIVIEELGVIGAGFILALVFFLI 60  
+ANS YAMSNGGWFG+GLGNSIEK GYLPEA TDFVFS+VIEELG+IGAGFILALVFFLI  
Sbjct: 269 LANSYYAMSNGGWFGQGLGNSIEKRGYLPEAQTDFVFSVIEELGLIGAGFILALVFFLI 328

40 Query: 61 LRIMHVGIAKADPFNSMIALGIGAMLLMQVFVNIGGISGLIPSTGVTFPFLSQGNSLLV 120  
LRIM+VGIAK+PFN+M+ALG+G M+LMQVFVNIGGISGLIPSTGVTFPFLSQGNSLLV  
Sbjct: 329 LRIMHVGIAKKNPFNMMALGVGGMMLMQVFVNIGGISGLIPSTGVTFPFLSQGNSLLV 388

45 Query: 121 LSVAGFVLNIDANEKKELIMKEAEQYK--PQEKNEKIINLDAFK 164  
LSVA+GFVLNIDA+EK++ I KEAE Y+ +++N K++N+ F+  
Sbjct: 389 LSVAVGFVLNIDASEKRDDIFKEAELSYRKDTRKENSKVNIKQFQ 434

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 50 Example 319

A DNA sequence (GBSx0348) was identified in *S.agalactiae* <SEQ ID 1029> which encodes the amino acid sequence <SEQ ID 1030>. This protein is predicted to be probable cell division protein ftsW (ftsW). Analysis of this protein sequence reveals the following:

Possible site: 34

55 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -9.77	Transmembrane	12 - 28 ( 7 - 37)
INTEGRAL	Likelihood = -7.22	Transmembrane	76 - 92 ( 74 - 97)
INTEGRAL	Likelihood = -6.53	Transmembrane	182 - 198 ( 178 - 201)
INTEGRAL	Likelihood = -4.62	Transmembrane	51 - 67 ( 46 - 69)
60 INTEGRAL	Likelihood = -2.87	Transmembrane	202 - 218 ( 202 - 218)

-404-

----- Final Results -----

5 bacterial membrane --- Certainty=0.4906(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9327> which encodes amino acid sequence <SEQ ID 9328> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:CAA44490 GB:X62621 ORF2 N-terminal [Lactococcus lactis]  
 Identities = 82/199 (41%), Positives = 122/199 (61%), Gaps = 9/199 (4%)

Query: 1 MKIDKRHLNYSILIPYLILSLGLIVIYSTTSATLIQLGANPFRSVINQGVFVAVSLVA 60  
 M ++K + LNYSILIPYLIL+ +G+++I+STT +Q G NP++ VINQ F +S++  
 15 Sbjct: 1 MNLNKNFLNYSILIPYLILAGIGIVMIFSTTVPDQLQKGLNPYKLVINQTAFLVLSIIM 60

Query: 61 IIFIYKLLNFKNSKVLTMVFLVEVFLLLIARF-----FTQEVNGAHGWIVIGPI-SF 113  
 I IY+LKL LKN K++ + +++ + L+ R T VNGA GWI I I +  
 20 Sbjct: 61 IAVIYRLKLRALKNRKMIGIIMVILILSLIFCRIMPSSFALTAPVNGARGWIHIPGIGTV 120

Query: 114 QPAEYLKVIIWYLAFTFARRQKKIEIYDYQALTGGRWLPRSLSDLKDWRFYSLFMIGLV 173  
 QPAE+ KV I+WYLA F+ +Q++IE D + KG+ L + L WR + ++ +  
 25 Sbjct: 121 QPAEFAKVFIWYLASVFSTKQEEIEKNDINEIFKGKTLTQKL--FGGWRLPVVAILLVD 178

Query: 174 IAQPDGLNGSIIVLTVIIM 192  
 + PDLGN II +IM  
 30 Sbjct: 179 LIMPDLGNTMIIGAVAILIM 197

There is also homology to SEQ ID 1028.

30 A related GBS gene <SEQ ID 8545> and protein <SEQ ID 8546> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 6  
 McG: Discrim Score: 15.18  
 35 GvH: Signal Score (-7.5): -3.58  
 Possible site: 34  
 >>> Seems to have an uncleavable N-term signal seq  
 ALOM program count: 5 value: -9.77 threshold: 0.0

INTEGRAL	Likelihood = -9.77	Transmembrane	12 - 28 ( 7 - 37)
INTEGRAL	Likelihood = -7.22	Transmembrane	76 - 92 ( 74 - 97)
INTEGRAL	Likelihood = -6.69	Transmembrane	210 - 226 ( 201 - 227)
INTEGRAL	Likelihood = -6.53	Transmembrane	182 - 198 ( 178 - 201)
INTEGRAL	Likelihood = -4.62	Transmembrane	51 - 67 ( 46 - 69)
PERIPHERAL	Likelihood = 1.32		116

modified ALOM score: 2.45  
 45 \*\*\* Reasoning Step: 3

----- Final Results -----

50 bacterial membrane --- Certainty=0.4906(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

55 ORF02700(301 - 876 of 1377)  
 EGAD|8615|8419(1 - 197 of 198) hypothetical protein in rpmg 3'region , fragment  
 {Lactococcus lactis} SP|P27174|YRG2\_LACLA HYPOTHETICAL PROTEIN IN RPMG 3'REGION (ORF2)  
 (FRAGMENT). GP|44069|emb|CAA44490.1||X62621 ORF2 N-terminal {Lactococcus lactis}  
 PIR|PC1134|PC1134 hypothetical protein 198 (rmpG 3' region) - Lactococcus lactis (fragment)  
 %Match = 15.1  
 60 %Identity = 42.3 %Similarity = 64.9  
 Matches = 82 Mismatches = 64 Conservative Sub.s = 44

```

87      117      147      177      207      237      267      297
KA*I*Y*I**L*LVLILFLLPFFINFL*IYLTGLND*NVPNSNIS*SFI FVISIVGGYXX*LIXXXIMHNGNFKY*R*K*Y

5       327      357      387      417      447      477      507      537
NMKIDKRHLNYSILIPYLILSILGLLIVIYSTTSATLIQLGANPFRRSVINQGVFWAVSLVAIIFIYKLKLNFLKNKVLT
|::||::| ||||| |||| :::||::| :| ||:: ||| | :|:: ||:| ||| ::
MNLNKNFNLYSILIPYLILAGIGIVMIFST'VDPDQLQKGLNPYKLVINQTAFVLLSTIMI AVIYRLKLRALKNRKMIG

10      10      20      30      40      50      60      70

567      585      609      636      666      696      726      756
MAVLVEVFLLLLIARF----FT--QE VNGAHGWIVIGPI-SFQPAEY LKVII VWYLAFTFARRQKKIEIYDYQALTKGRWL
: :: :: |: | | ||| ||| | | : |||: || |: ||| |: :|::| | : |: |
IIMVILILSLIFCRIMPSSFALTAPVNGARGWIHIPGIGTVQPAEF AKVFIIWYLASVFSTQKEEIEKNDINEIFKGKTL

15      90      100      110      120      130      140      150

786      816      846      876      906      936      966      996
PRSLSDLKDWRFSYSLFMIGLVIAQPDLGN GSIIVLTVIIMYCISGIGYRWFSALLGLLIVVGSTLFIGTIAVVGVMKAV
: | : ||: : :: : |||| || :||
TQKL--FGGWRLPVVAILLVDLIMPDLGNTMIIGAVALIMI

20     170     180     190

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 320

25 A DNA sequence (GBSx0349) was identified in *S.agalactiae* <SEQ ID 1031> which encodes the amino acid sequence <SEQ ID 1032>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence
```

```

30      ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.3665(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

35 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1033> which encodes the amino acid sequence <SEO ID 1034>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence
```

```

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2373(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

Identities = 35/41 (85%), Positives = 37/41 (89%)

50      Query: 1 MEKEAKQIIDLKRNLFKIDVRAQKDEEKVFMRTACCYSPPY 41  
             +EKEAKQ+IDLKRNLFKIDVRAQKDEEKVFMRTAC S Y  
             Subject: 1 LEKEAKQMIDLKRNLFKIDVRAQKDEEKVFMRTACRQSRVY 41

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-406-

**Example 321**

A DNA sequence (GBSx0351) was identified in *Sagalactiae* <SEQ ID 1037> which encodes the amino acid sequence <SEQ ID 1038>. Analysis of this protein sequence reveals the following:

Possible site: 49

5 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.65 Transmembrane 78 - 94 ( 78 - 95)  
 INTEGRAL Likelihood = -1.33 Transmembrane 421 - 437 ( 420 - 437)

----- Final Results -----

10 bacterial membrane --- Certainty=0.1659(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15 >GP:CAA00827 GB:A09073 phosphoenol pyruvate carboxylase  
 [Corynebacterium glutamicum]  
 Identities = 335/958 (34%), Positives = 539/958 (55%), Gaps = 80/958 (8%)

20 Query: 22 EIITEEVGLLKQLLDEATQKLIGSESFDKIE--KIVSLSLTD---DYTGLKETISALSNE 76  
 + + +++ L Q+L E + G E ++ +E ++ S + + L + ++  
 Sbjct: 3 DFLRDDIRFLGQILGEVIAEQEGQEVYELVEQARLTSFDIAKGNAEMDSLQVDFDGITPA 62

25 Query: 77 EMVIVSRYFSILPLLINISEDVDLAYEINYKNNLNQDYLGLKSLT----TIDVV----- 125  
 + ++R FS LL N++ED+ Y L + L T T+D  
 Sbjct: 63 KATPIARAFSHFALLANLAEDL-----YDEELREQALDAGDTPPDSTLDATWLKLNEG 115

30 Query: 126 -AGHENAKDILEHVNVVPVLTHTPTQVQRKTVLELTSKIHDLLRKYRDVKAGIVNQ---- 180  
 G E D+L + V PVLTAHPT+ +R+TV + I +R+ +++  
 Sbjct: 116 NVGAEAVADVLRNAEVAPVLTHTPTETRRRTVFDQKWITTHMRERHALQSAPTARTQS 175

35 Query: 181 --EKWYADLRRYIGIIMQTDITREKKLVKNEITNVMYYNRSLIKAVTKLTAEYKALAA 238  
 ++ ++RR I I+ QT IR + +++EI + YY SL++ + ++ +  
 Sbjct: 176 KLDEIEKNIRRRITILWQTALIRVARPRIEDEIEVGLRYKLSLLEEIPRINRDVAVELR 235

40 Query: 239 KK---GIHLENPKPLTM-GMWIGGDRDGNPFVTAETLRLSAMVQSEVIINHYIEQLNELY 294  
 ++ G+ L KP+ G WIGD DGNP+VTAET+ S +E ++ +Y QL+ L  
 Sbjct: 236 ERFGEVPL---KPVVKPGSWIGGDHGNPYVTAETVEYSTHRAAETVLKYYARQLHSLE 292

45 Query: 295 RNMSLSINLTVSPVLTLANQSQDNSVYRENEPYRKAFNFIQDKLVQTLNLKVGSSPK 354  
 +SLS + +V+P+L+ LA+ ++ R +EPYR+A + ++ +++ T  
 Sbjct: 293 HELSLSDRMNKVTPQLLALADAGNDVPSRVDEPYRRVHGVGRILAT----- 341

50 Query: 355 EKFSVRQESSDIVGRYIKSHIAQVASDIQTEELPAYATAEEFKQDLLLVKQSLVQYGQDS 414  
 +++++G + + YA+ EEF D L + SL +  
 Sbjct: 342 -----TABELIGE-----DAVEGVWFKVFTFYASPEEFLNDALTIDHSLRESKDV 386

55 Query: 415 LVDGELACLIQAVDIFGFYLATIDMRQDSSINEACVAELLKSANIVDDYSSLSEEEKCQL 474  
 + D L+ LI A++ FGF L +D+RQ+S E + EL + A + +Y LSE EK ++  
 Sbjct: 387 IADDRLSVLISAIESFGFNLYALDLRQNSSESYEDVLTELFERAQVTANYRELSEAEKLEV 446

60 Query: 475 LLKELTEDPRTLSTHAPKSELLQKELAIQTARELKDQLGEDIINQHIISHTESVSDMF 534  
 LLKEL + SE+ +EL IF+TA E + G ++ IIS SV+D+  
 Sbjct: 447 LLKELRSPRPLPHGSDEYSEVTPRELGIPTASEAVKKFGPRMVPHCIISSMASSVTDVL 506

55 Query: 535 ELAIMLKEVGLIDAN---QARIQIVPLFETIEDLDNSRDIMTQYLHYELVKKWIATNWN 590  
 E ++LKE GLI AN + + ++PLFETIEDL I+ + +L + ++ +N  
 Sbjct: 507 EPMVLLKEFGLIAANGDNPRGTVDVPLFETIEDLQAGAGILDELWKIDLYRNYLLQRDN 566

60 Query: 591 YQEIMLGYSDSNKGGLYSSGWTLKYKAQNELTKIGEENGIKITFFHGRGGTVGRGGGPSY 650  
 QE+MLGYSDSNKGGY S+ W LY A+ +L ++ G+K+ FHGRGGTVGRGGGPSY  
 Sbjct: 567 VQEVMLGYSDSNKGGYFSANWALYDAELQLVELCRSAGVKLRLFHGRGGTVGRGGGPSY 626

Query: 651 EAITSQPFSGIKDRIRLTEQGEIENKYGNQDAAYYNLEMLISASIDRMVTRMITNPNEI 710  
 +AI +QP G+++ +R+TEQGEII KYGN + A NLE L+SA+++ + + +E+



-407-

Sbjct: 627 DALLAQPRGAVQGSVRITEQGEIISAKYGNPETARRNLEALVSATLE----ASLLDVSEL 682

Query: 711 DNFRETMDGIVSESNAV----YRNLVFDNPFYFYDYFFEASPIKEVSSLNIGSRPAARKTI 766  
 + + D I+SE + + Y +LV ++ F DYF +++P++E+ SLNIGSRP++RK

5 Sbjct: 683 TDHQRAYD-IMSEISELSLKKYASLVHEDQGFIDYFTQSTPLQEIGSLNIGSRPSSRKQT 741

Query: 767 TEISGLRAIPWVFSWSQNRIMFPGWYGVGSFAKHF----EQDEANLAKLQTMYPKWPFFN 823  
 + + LRAIPWV SWSQ+R+M PGW+GVG+A + +I EQ +A+LQT+ + WFFF

10 Sbjct: 742 SSVEDLRAIPWVLSWSQSRVMLPGWFGVGTALEQWIGEGEQATQRIAEQLTLNESWPFPT 801

Query: 824 SLLSNVDMVLSKSNMNIALQYAQLAGSKEVRD-VFNIIILNEWQLTKDMILAEQHDNLE 882  
 S+L N+ V+SK+ + +A YA L EV + V+++I E+ LTK M I D+LL+

Sbjct: 802 SVLDNMAQVMSKAELRLAKLYADLIPDTEVAERVYSVIREEYFLTKKMFVCITGSDDLLD 861

15 Query: 883 ENPMLHASLDYRLPYFNVLYVQIELIKRLRSNQLDEDEYKLIHITINGIATGLRNSG 940  
 +NP+L S+ R PY LN +Q+E+++R R E + I +T+NG++T LRNSG

Sbjct: 862 DNPLLRASVQRRYPYLLPLNVIQVEMMRRYRKGDQSEQVSRNIQLTMNGLSTALRNSG 919

A related GBS nucleic acid sequence <SEQ ID 10961> which encodes amino acid sequence <SEQ ID 10962> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1039> which encodes the amino acid sequence <SEQ ID 1040>. Analysis of this protein sequence reveals the following:

Possible site: 40  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1613(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 659/927 (71%), Positives = 779/927 (83%), Gaps = 11/927 (1%)

35 Query: 14 KLESSNKEIITEEVGLLKQLLDEATQKLIGSESFDKIEKIVSLSLTDDYTGLKETISAL 73  
 KLESS+N++II EEV LLK++L+ T+++IG ++F IE I+ LS DY L++ ++ +

Sbjct: 5 KLESSNNQDIIAEVALLKEMLNITRRMIGDGAFTVIESIMVLSEKQDYIELEKVVANI 64

Query: 74 SNEEMVIVSRYSILPLLINISEDVDLAYEINYKNNLNQDYLGLKLTITIDVVAGHENAKD 133  
 SN+EM ++SRYSILPLLINISEDVDLAYEINY+NN + DYLGLK+ TI +AG +N KD

40 Sbjct: 65 SNQEMEIVISRYFSILPLLINISEDVDLAYEINYQNNTDTDYLGKLALTIKDLAGKDKNGKD 124

Query: 134 ILEHVNVPVLTAHPTQVQRKTVLELTSKIHDLRLRYRDKAGIVNQEKWYADLRRYIGI 193  
 ILE VNVPVLTAHPTQVQRKT+LELT+ IH LLRKYRD KAG++N EKW +L RYI +

45 Sbjct: 125 ILEQVNVPVLTAHPTQVQRKTILELTTHIHKLLRKYRDAKAGVINLEKWRQELYRYIEM 184

Query: 194 IMQTDITIREKKLVKNEITNMVEYNNRSLIKAVTKLTAEYKALAAKKGHIHLENPKPLTMG 253  
 IMQTD IREKKL+VKNEI NVM+YY+ SLI+AVTKLT EYK LA K G+ L+NPKP+TMG

Sbjct: 185 IMQTDIIREKKLVKNEIKVMQYDGSLSLQAVTKLTTEYKNAQKHGLELDNPKPITMG 244

50 Query: 254 MWIGGDRDGNPFVTAETLRLSAMVQSEVIINHIEQLNELYRNMSLSINLTVSPSELVTL 313  
 MWIGGDRDGNPFVTAETL LSA VQSEVI+N+YI++L LYR SLS L + + E+ L

Sbjct: 245 MWIGGDRDGNPFVTAETLCLSATVQSEVILNYYIDELAALYRTFSLSSSTLVQPNSEVERL 304

55 Query: 314 ANQSQDNSVYRENEPYRKAFNFIQDKLVQTLNLKVGSSPKKFKVSRQESSDIVGRYIKS 373  
 A+ SQD S+YR NEPYR+AF++IQ +L QT + L + + SS + S

Sbjct: 305 ASLSQDQSIYRGNEPYRRAFHYIQSRLKQTQIQLT-----NQPAASMSSSVGLNTSAWS 358

Query: 374 HIAQVASDIQTEELPAYATAEEFKQDLLLVKQSLVQYQDQSLVDGELACLIQAVDIFGFY 433  
 A + + I AY + +FK DL ++QSL+ G +L++G+L ++QAVDIFGF+

60 Sbjct: 359 SPASLENPIL-----AYDSPVDFKADLKAIEQSLLDNGNSALIEGDLREVMQAVDIFGFF 413

Query: 434 LATIDMRQDSSINEACVAELLKSANIVDDYSSLSEEEKCQLLLKELTEDPRTLSSTHAPK 493  
 LA+IDMRQDSS+ EACVAELLK ANIVDDYSSLSE EKC +LL++L E+PRTLS K

-408-

- Sbjct: 414 LASIDMRQDSSVQEACVAELLKGANIVDDYSSLSETEKCDVLLQQLMEEPRTLSSAAVAK 473
- Query: 494 SELLQKELAI FQTARELKDQLGEDI INQHIISHTESVSDMFELAIMLKEVGLIDANQARI 553  
S+LL+KELAI+ TARELKD+LGE++I QHIISHTESVSDMFELAIMLKEVGL+D +AR+
- 5 Sbjct: 474 SDLLEKELAIYTTARELKD KLGEV I KQHIISHTESVSDMFELAIMLKEVGLVDQQRARV 533
- Query: 554 QIVPLFETIEDLDNSRDIMTQYLHYELVKKWIATNNNYQEIMLGYSDSNKDGGYLSSGWT 613  
QIVPLFETIEDLDN+RDIM YL +++VK WIATN NYQEIMLGYSDSNKDGGYL+SGWT
- 10 Sbjct: 534 QIVPLFETIEDLDNARDIMAAYLSHDIVKSWIATNRNYQEIMLGYSDSNKDGGYLASGWT 593
- Query: 614 LYKAQNELTKIGEENG I KITFFHGRGGTVGRGGGPSYEAITSQPFSGIKDRIRLTEQGEI 673  
LYKAQNELT IGEE+G+KITFFHGRGGTVGRGGGPSY+AITSQPFSGIKDRIRLTEQGEI
- Sbjct: 594 LYKAQNELTAIGEEHGVKITFFHGRGGTVGRGGGPSYDAITSQPFSGIKDRIRLTEQGEI 653
- 15 Query: 674 IENKYGNDAAAYNLEMLISASIDRMVTRMITNPNEIDNFRETMDGIVSESNAVYRNLFV 733  
IENKYGND AYY+LEMLISASI+RMVT+MIT+PNEID+FRE MD IV++SN +YR LVF
- Sbjct: 654 IENKYGNDVAYYHLEMLISASINRMVTQMIDPNEIDSFREIMDSIVADSNIIYRKLVF 713
- 20 Query: 734 DNPYFYDYFFEASPIKEVSSLNIGSRPAARKTITEISGLRAIPWVFSWSQNRIMFPGWYG 793  
DNP+FYDYFFEASPIKEVSSLNIGSRPAARKTITEI+GLRAIPWVFSWSQNRIMFPGWYG
- Sbjct: 714 DNPYFYDYFFEASPIKEVSSLNIGSRPAARKTITEITGLRAIPWVFSWSQNRIMFPGWYG 773
- Query: 794 VGSFAFKHFIEQDEANLAKLQTMYPKWPFFNSLLSNVDMVLKSNMNIALQYAQLAGSKEV 853  
VGSFAFK +I++ + NL +LQ MYQ WPF+SLLSNVDMVLKSNMNIA QYAQLA ++V
- 25 Sbjct: 774 VGSFAKRYIDRAQGNLERLQHMYPKWPFFHSLLSNVDMVLKSNMNIAFQYAQLAERQDV 833
- Query: 854 RDVFNIILNEWQLTKDMILAIEQHDNLL ENPMLHASLDYRLPYFNVNLNYQIELIKRLR 913  
RDVF IL+EWQLTK++ILAI+ HD+LLE+NP L SL RLPYFNVNLNY+QIELIKR R
- 30 Sbjct: 834 RDVFYEILDEWQLTKNVILAIQHDHLLDNPSLKHSLKSRLPYFNVNLNYQIELIKRWR 893
- Query: 914 SNQLDEDEYKLIHITINGIATGLRNSG 940  
+NQLDE+ EKLIH TINGIATGLRNSG
- Sbjct: 894 NNQLDENDEKLIHTTINGIATGLRNSG 920
- 35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 322**

- A DNA sequence (GBSx0352) was identified in *S.agalactiae* <SEQ ID 1041> which encodes the amino acid sequence <SEQ ID 1042>. This protein is predicted to be *Bacillus licheniformis* Pz-peptidase
- 40 homologue (pepF). Analysis of this protein sequence reveals the following:

Possible site: 61  
>>> Seems to have no N-terminal signal sequence

- 45 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.3012(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1043> which encodes the amino acid sequence <SEQ ID 1044>. Analysis of this protein sequence reveals the following:
- 50

Possible site: 55  
>>> Seems to have no N-terminal signal sequence

- 55 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.3137(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

-409-

Identities = 512/593 (86%), Positives = 564/593 (94%)

Query: 1 MKLKKRSEFPENELWDLTALYKDRQDFLLAIEKALEDIKVFKNYEGKLNVEDFTSALM 60  
 Sbjct: 26 MELKKRSEFPENELWDLTALYKDRQDFLLAIEKALQDIDLFRKNYEGRLTSVDDFTQALI 85

Query: 61 EIEHIYIQMSHIDTYAFMPQTTDFSNEEFAQISQAGSDFATKANVLLSFFNTALANADIK 120  
 EIEHIYIQMSHI TYAFMPQTTDFS+E FAQI+QAG DF TKA+V LSFF+TALANAD+  
 Sbjct: 86 EIEHIYIQMSHIGTYAFMPQTTDFSDESFAQIAQAGDDFMTKASVALSFFDTALANADLD 145

Query: 121 ILDSLENNPHFKATIRQAKIQKHLLSPEVEKALTNLNEVLNTPYDIYTKMRAGDFDMED 180  
 +LD+LE NP+F A IR AKIQK+HLLSP+VEKAL NL EV+N PYDIYTKMRAGDFDM+D  
 Sbjct: 146 VLDTLEKNPYFSAAIRMAKIQKEHLLSPDVEKALANLREVINAPYDIYTKMRAGDFDMDD 205

Query: 181 FEVDGKTYKNSFVYENYFQNHENAEIREKSFRSFSKGLRKHQNAAYLAKVKSEKLI 240  
 FEVDGKTYKNSFV+YEN++QNHENAEIREK+FRSFSKGLRKHQNAAYLAKVKSEKLI+  
 Sbjct: 206 FEVDGKTYKNSFVSFVYENYFQNHENAEIREKAFRSFSKGLRKHQNTAAAYLAKVKSEKLL 265

Query: 241 ADMRGYDSVFDYLLSEQEVDRSMFDRQIDLIMDEFGPVAQRFLKHIADVNGIEKMTFADW 300  
 ADM+GY SVFDYLL+EQEVDRS+FDRQIDLIM EFGPVAQ+FLKH+A VNG+EKMTFADW  
 Sbjct: 266 ADMKGYASVFDYLLAEQEVDRSLFDRQIDLIMTEFGPVAQKFLKHVAQVNGLEKMTFADW 325

Query: 301 KLDIDNELNPEVSINDAYDLVMKSVAPLGKEYSQEVERYQKERWVDFANANKDSGGYAA 360  
 KLDIDN+LNPEVSI+ AYDLVMKS+APLG+EY++E+ERYQ ERWVDFANANKDSGGYAA  
 Sbjct: 326 KLDIDNDLNPEVSIDGAYDLVMKSLAPLGQEYTKIERYQTERWVDFANANKDSGGYAA 385

Query: 361 DPYKVHPYVLMWSWTGRMSDVYTLIHEIGHSGQFIFSDNHQSFFNTHMSTYYYEAPSTFNE 420  
 DPYKVHPYVLMWSWTGRMSDVYTLIHEIGHSGQFIFSDNHQS+FNTNTHMSTYYYEAPSTFNE  
 Sbjct: 386 DPYKVHPYVLMWSWTGRMSDVYTLIHEIGHSGQFIFSDNHQSYFNTNTHMSTYYYEAPSTFNE 445

Query: 421 LLLSDYLENQFDTRQKRFALAHRLTDYFHNFIHLLLEAAFQRKVYTLIEEGGTFGAEQ 480  
 L+LSDYLE+QFD RQKRFALAHRLTDYFHNFIHLLLEAAFQRKVYTLIEEGGTFGA+Q  
 Sbjct: 446 LMLSDYLEHQFDDPRQKRFALAHRLTDYFHNFIHLLLEAAFQRKVYTLIEEGGTFGADQ 505

Query: 481 LNAIMKEVLTQFWGDAIEIDDDAALTWMRQAHYMGLYSYTYSAGLVISTAGYLNKKNP 540  
 LNA+MKEVLT FWGDA++IDDDAALTWMRQAHYMGLYSYTYSAGLVISTAGYLNK+NP  
 Sbjct: 506 LNAMMKEVLTDFWGDVIDDDAALTWMRQAHYMGLYSYTYSAGLVISTAGYLNKKNP 565

Query: 541 NGAKEWLFLKSGGSRTPLLETALLISADISTDKPLRDTINFLSNTVDQIINYS 593  
 NGAKEWL FLKSGGSRTPL+TA+LI ADI+T+KPLRDTI FLS+TVDQII+Y+  
 Sbjct: 566 NGAKEWLFLKSGGSRTPLDTAMLIGADIATEKPLRDTIQFLSNTVDQIISYT 618

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 323

A DNA sequence (GBSx0353) was identified in *Sagalactiae* <SEQ ID 1045> which encodes the amino acid sequence <SEQ ID 1046>. Analysis of this protein sequence reveals the following:

Possible site: 19  
 >>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1047> which encodes the amino acid sequence <SEQ ID 1048>. Analysis of this protein sequence reveals the following:

Possible site: 19

-410-

&gt;&gt;&gt; May be a lipoprotein

----- Final Results -----

5                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

10           Identities = 72/127 (56%), Positives = 85/127 (66%)

Query: 1   MKKYIKLFLLTVFATTLVACGQPSTSNKTTTSSSTLEVKGKVELVVKEDTNVLSEKVVYHKG 60  
           + K K L + A LVAC Q + +TT S V LVVKEDTN + EKV + KG  
 Sbjct: 1   VNKRFKTGFLALVAMLLVACSQGTKQIQTTSPVPKADHHVRLVVKEDTNTVDEKVSFGKG 60

15           Query: 61   DTVLDVLKANYKVKEKDGFIITSIDGISQDETKGLYWMFKVNNKLAPKAANQIKVKKNDKI 120  
                   DTVL+VLK NY+VKEKDGFIIT+IDGI QD YW+FKVN K+A K A+QI VK D I  
           Sbjct: 61   DTVLEVLKDNYEKVDGFIITAIDGIEQDTKANKYWLFKVNGKMADKGADQITVKDGDSI 120

20           Query: 121 EFYQEVY 127  
                   EFYQEV+  
           Sbjct: 121 EFYQEVF 127

25           SEQ ID 1046 (GBS185) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 28 (lane 6; MW 15.7kDa).

GBS185-His was purified as shown in Figure 199, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 324

30           A DNA sequence (GBSx0354) was identified in *S.agalactiae* <SEQ ID 1049> which encodes the amino acid sequence <SEQ ID 1050>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

35           INTEGRAL   Likelihood = -4.46   Transmembrane   75 - 91 ( 67 - 94)  
                   INTEGRAL   Likelihood = -4.41   Transmembrane   33 - 49 ( 30 - 49)  
                   INTEGRAL   Likelihood = -2.60   Transmembrane   53 - 69 ( 52 - 70)  
                   INTEGRAL   Likelihood = -1.38   Transmembrane   108 - 124 ( 106 - 124)  
                   INTEGRAL   Likelihood = -0.06   Transmembrane   149 - 165 ( 149 - 165)

40           ----- Final Results -----

                  bacterial membrane --- Certainty=0.2784 (Affirmative) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45           A related GBS nucleic acid sequence <SEQ ID 9731> which encodes amino acid sequence <SEQ ID 9732> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10929> which encodes amino acid sequence <SEQ ID 10930> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

50           A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1051> which encodes the amino acid sequence <SEQ ID 1052>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have a cleavable N-term signal seq.

-411-

```

INTEGRAL    Likelihood = -7.96    Transmembrane    50 - 66 ( 49 - 71)
INTEGRAL    Likelihood = -5.73    Transmembrane    101 - 117 ( 99 - 124)
INTEGRAL    Likelihood = -4.41    Transmembrane    141 - 157 ( 139 - 159)
INTEGRAL    Likelihood = -4.25    Transmembrane    73 - 89 ( 67 - 92)

```

----- Final Results -----

```

bacterial membrane --- Certainty=0.4185(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

Identities = 82/163 (50%), Positives = 120/163 (73%), Gaps = 3/163 (1%)

```

Query: 10  LTRVAILSALCVVLR YAFAPLPNIQPITAIFLITVVLFDLKEGVATVTITMLVSSFLMGF 69
          ++R+AI+SALCVVLR  F+ LPN+QP+TA  L  ++ F L E V  + + + +S+FL+GF
Sbjct: 6   MSRIAIMSALCVVLRMVFSLEPNVQPVTAFLLSYLLYFGLAEAVLVMMLCLFLSAFLLGF 65

Query: 70  GPWVFLQIISFTLILCLWKFLIYPLTKAVCFGKITEVVLQTFAGGLGVVYGVIIIDTCFA 129
          GPWVF Q+ F L+L LW+F++YPL++  F K  ++ Q F  G++YGV+IDTCFA
Sbjct: 66  GPWVFWQVTCFVLVLLWRFVLYPLSQQ--FPKY-QLGCQAFVLVLCGLLYGVLIIDTCFA 122

Query: 130 WLYHMPWWTYVVLAGLSFNMAHALSTCLFYPLLLPILRRFRNEK 172
          +LY MPWW+YVLAG+ FN+AHALST +F+P+++ + RR  E+
Sbjct: 123 YLYSMPWWSYVLAGMPFNIAHALSTLVFFPVVMMLEFRRLIGE 165

```

A related GBS gene <SEQ ID 8549> and protein <SEQ ID 8550> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1  Crend: 10
McG: Discrim Score: 6.79
GvH: Signal Score (-7.5): -0.91
    Possible site: 28
>>> Seems to have a cleavable N-term signal seq.
ALOM program  count: 3 value: -4.46 threshold: 0.0
INTEGRAL    Likelihood = -4.46    Transmembrane    35 - 51 ( 29 - 54)
INTEGRAL    Likelihood = -1.38    Transmembrane    68 - 84 ( 66 - 84)
INTEGRAL    Likelihood = -0.06    Transmembrane    109 - 125 ( 109 - 125)
PERIPHERAL  Likelihood = 7.53      88
modified ALOM score: 1.39

*** Reasoning Step: 3

----- Final Results -----
bacterial membrane --- Certainty=0.2784(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

ORF01220(421 - 552 of 1002)
GP|9950155|gb|AAG07353.1|AE004814_8|AE004814(16 - 56 of 69) hypothetical protein
{Pseudomonas aeruginosa}
%Match = 3.2
%Identity = 39.5 %Similarity = 60.5
Matches = 17 Mismatches = 15 Conservative Sub.s = 9

222      252      282      312      342      372      402      432
STLTKLTRVAILSALCVVLR YAFAPLPNIQPITAIFLITVVLFDLKEGVATVTITMLVSSFLMGFGPWVFLQIISFTLIL
                                                                |:::
                                                                MDPELFEEWMMTGLVTVLI
                                                                10

462      492      522      552      582      612      642      672
CLWKFLIYPLTKAVCFGKITEVVLQTFAGGLGVVYGVIIIDTCFAWLYHMPWWTYVVLAGLSFNMAHALSTCLFYPLLLPI

```

-412-

```

      : |:: | |      ||  ::|  ||| ||| | ||
LFMAFIVWDLAKSKAGKFGTLIL--FFALGLGV-LGFIKGLVIGSLEGAGM
      30      40      50      60

```

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 325

- 10 A DNA sequence (GBSx0355) was identified in *S.agalactiae* <SEQ ID 1053> which encodes the amino acid sequence <SEQ ID 1054>. This protein is predicted to be endolysin. Analysis of this protein sequence reveals the following:

```

Possible site: 28
>>> Seems to have a cleavable N-term signal seq.

```

- ```

----- Final Results -----
15      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

- ```

20 >GP:CAA72266 GB:Y11477 endolysin [Bacteriophage Bastille]
    Identities = 64/210 (30%), Positives = 95/210 (44%), Gaps = 15/210 (7%)

Query: 66 KPIIDVSGWQLPKIEDYDTLSKNISGVVIRVFGGSKISKTNNAAYTTGIDKSFKTHIKEF 125
      K I+D+S      +ID+DT      +S + R      G + + +N      +D+ +KT +
25 Sbjct: 12 KTIVDISHHNA--DIDFDTAKNVYSMFIARTGDGHRYN--SNGELQGVVDRKYKTFVANM 67

Query: 126 QKRNIPIVAVYSYALGSSVKEMKEEAQIFYKNAAPYKPTFYWIDVEEETMSNMNKGVAQFR 185
      + R IP      Y +      S V      K+EA+ F+ N      T + D E T NM + +Q F
30 Sbjct: 68 KARGIPFGNYMFNRFSGVASAKQEAFFW-NYGDKDATVWVCDAEVSTAPNMKECIQVFI 126

Query: 186 KELKRLGAKNVGIYIGTYFMTEQGISVKGFDAVWIPTYGSDSGYYEAAPOTELKYDLHQY 245
      LK LGAK VG+YIG +      E G      D WIP YG+      +      DL Q+
35 Sbjct: 127 DRLKELGAKKVGLYIGHHKYQEFGGKDVNCDFTWIPRYGNKPAF-----ACDLWQW 177

Query: 246 TSQGYLPGFNQPLDLNQIAVNKDKKKTYEK 275
      T G + G + D+N + +K      EK
40 Sbjct: 178 TEYGNIAIGIK-CDINVLYGDKPMSFFTEK 206

```

- 40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1055> which encodes the amino acid sequence <SEQ ID 1056>. Analysis of this protein sequence reveals the following:

```

Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL      Likelihood =-16.98      Transmembrane      8 - 24 (      3 - 28)

```

- ```

45 ----- Final Results -----
      bacterial membrane --- Certainty=0.7793(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

- 50 An alignment of the GAS and GBS proteins is shown below:

```

Identities = 198/278 (71%), Positives = 235/278 (84%)

```

- ```

Query: 1 MRRRIKPIVVAVFFSLFGLLLIIGHLHSTNTLKKELVEAKKTIPSVKASKVPQKSTSSKD 60
      MRR+IKPIVV VFF L ++LIIG      + + +KE+ +AK IP      ++      K+++S+
55 Sbjct: 1 MRRKIKPIVVLVFFILLAMVLIIGKQANHAQKEVEDAKSHIPIATSNPGKAKTSTSET 60

Query: 61 KEFVLKPIIDVSGWQLPKIEDYDTLSKNISGVVIRVFGGSKISKTNNAAYTTGIDKSFKT 120
      ++F+L PI+DVSGWQLP+EIDYDTLS++ISG ++RV+GGS+I+      NNAA+TTGIDKSFKT

```



-414-

PAFACDLWQWTEYGNIAIGIGK-CDINVLYGDKPMSFFTEKEGAKETLVPALNKVVTVYEVGTNLIPETIQDKLAFLLGYEARL  
 180 190 200 210 220 230 240

SEQ ID 8552 (GBS206) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 6; MW 31.7kDa).

GBS206-His was purified as shown in Figure 206, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 326

A DNA sequence (GBSx0356) was identified in *S.agalactiae* <SEQ ID 1057> which encodes the amino acid sequence <SEQ ID 1058>. Analysis of this protein sequence reveals the following:

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.44 Transmembrane 183 - 199 ( 183 - 200)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1574(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9729> which encodes amino acid sequence <SEQ ID 9730> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG20117 GB:AE005090 NADH dehydrogenase/oxidoreductase-like  
 protein; Nola [Halobacterium sp. NRC-1]  
 Identities = 38/156 (24%), Positives = 83/156 (52%), Gaps = 13/156 (8%)  
 Query: 19 TMEILIAGSGFLGKQIIKAALT'KGHKVAYLSRHEGKGDIFKDPRLTYIRGDITEADKIH 78  
 +M++L+ GG+GF+G + + +GH V +R + D +T I GD+T + +  
 Sbjct: 8 SMDVLVTGGTGFIGTHLCRELDDRGHDVTAFAREPADALPAD--VTRIVGDVTVKETVA 65  
 Query: 79 LEDRTFDILIDCIGA---IKPNQLD---ELNVKATQKAVALCHKNQIPKLVYISA----- 127  
 D +++ + KP+ D ++++ T+ VA + + ++ +SA  
 Sbjct: 66 NAIDGHDVAVNLVALSPLFKPSGGDSRHLDVHLGGTENVVAAASEAGVEYIILQSLADAD 125  
 Query: 128 NSGYSAYIKSKRAEQIIKASGLDYLFVRPGLMYGE 163  
 +G +AY+++K +AE+ +++S L + VRP +++G+  
 Sbjct: 126 PTGPTAYLRAKGRABEAVRSSLHHTIVRPSVVFGD 161

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8553> and protein <SEQ ID 8554> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 5  
 McG: Discrim Score: -7.99  
 GvH: Signal Score (-7.5): -6.34  
 Possible site: 41  
 >>> Seems to have no N-terminal signal sequence  
 ALOM program count: 1 value: -1.44 threshold: 0.0  
 INTEGRAL Likelihood = -1.44 Transmembrane 183 - 199 ( 183 - 200)  
 PERIPHERAL Likelihood = 4.29 20  
 modified ALOM score: 0.79

\*\*\* Reasoning Step: 3



-415-

----- Final Results -----

bacterial membrane --- Certainty=0.1574(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

RGD motif 68-70

The protein has homology with the following sequences in the databases:

32.5/54.4% over 274aa

Schizosaccharomyces

pombe

GP|3395590|emb|CAA20132.1||AL031179(1 - 275 of 276) hypothetical protein

PIR|T41177|T41177 hypothetical protein SPCC1840.09 - fission yeast Insert characterized

ORF01216(358 - 990 of 1272)

GP|3395590|emb|CAA20132.1||AL031179(1 - 275 of 276) hypothetical protein  
 {Schizosaccharomyces pombe} PIR|T41177|T41177 hypothetical protein SPCC1840.09 - fission  
 yeast (Schizosaccharomyces pombe)

%Match = 7.3

%Identity = 32.4 %Similarity = 54.3

Matches = 71 Mismatches = 88 Conservative Sub.s = 48

144 174 204 234 264 294 324 354  
 \*L\*\*ISTDS\*K\*A\*IPFQGIMIINIATVLFGLN\*KFYK\*LNMKCPDVM\*NTTVVRY\*TITLTHRIKISILNLQNEGEG

384 414 444 474 504 534 564  
 TMEILIAGSGFLGKQIIKAALTKGHKVAYLSRHEGKGDIFKDPRLTYIRGDITEADKIHLEDRTFDILIDCIGAI----  
 |:|:: ||| ||| | | |:|::| :|| | |:| : : : :| | : : :| | :  
 MKIVVLGGSGFLGHNICKLATAKGYEVSVSRGAGGLHNKEPWWDDVEWETLDAQK--DPNSLLPVLRDASAVVNSVG  
 10 20 30 40 50 60 70

585 615 648 678  
 -----KPNQLDELNVKATQKAV-----ALCHKNQIPKLVIYIS  
 || : : | | :| : | :|  
 ILMENNYKKILQNPRGPVSHLINSLSNMFKTGONPLAPKPEEAKQSKNKVTFEAINRDLAIETAKIAAKANVPVYCYVS  
 90 100 110 120 130 140 150

699 726 753 783 810 840 846 876  
 ANS---GYSA-YIKSKRAE-QIIKASGLDYLFVRPGLMYG-EERPLSIFQAKCIKLFSLH-----PFLGIVVQKVF  
 |:: | |||:|:| | | | :|:|:|:| :|:|:| : | : ||| : :  
 AHAAPGLDPRYIKTKREAEREISKISNLRISIFLRPGFMYNFNDRPFTGALASLFTVSSSINRATSGALNFLGTASAEPL  
 170 180 190 200 210 220 230

930 960 990 1020 1050 1080 1110  
 PTK-VVIVA-EAIVTTLRKPTQKILSIEELNNK\*FIKKATVNSSFYSFTFPKSFS\*VFFLSLLTAI\*FKSSG\*LXPGR\*  
 |:: | : | || | | : | : : :| | :|:  
 PSEEVALAALEAISDPVSKGPVE-ISELKSMARK-FKQKSL  
 250 260 270

SEQ ID 8554 (GBS303) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 5; MW 28.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 5; MW 53.2kDa).

The GBS303-GST fusion product was purified (Figure 207, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 275), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 327**

A DNA sequence (GBSx0357) was identified in *S.agalactiae* <SEQ ID 1059> which encodes the amino acid sequence <SEQ ID 1060>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 49
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2850(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAC36853 GB:L23802 regulatory protein [Enterococcus faecalis]
Identities = 61/164 (37%), Positives = 96/164 (58%), Gaps = 13/164 (7%)

15  Query: 1  MSKKNKIKKTLVDQILDKAKIEH-----DSLQLDALQGDLPNGIQKQDIFKTLALI 51
      M+KK  +KT  +++++ K+ +          D L  +++  L  GI+K  IFKTL  +
      Sbjct: 1  MAKKKTKQKTNAMRMVEQHKKVPYKEYEFAWSEDHLSAESVAESL--GLEKGRIFKTLVTV 58

20  Query: 52  GDKTGPIIGILPLTEHLSEKKLAKISGNKKVQMIPQKDLQKITGYIHGANNPIGIRQKH 111
      G+KTGP++ ++P  + L  KKLAK SGNKKV+M+  KDL+  TGYI G  +P G+  K
      Sbjct: 59  GNKTGPVVAVIPGNQELDLKKLAKASGNKKVEMHLKDLKLEATGTGYIRGGCSPTGM--KKQ 116

25  Query: 112 YPIFIDTIALEKQELIVSAGEIGRSIRINSEVLADFVNAKFADI 155
      +P ++  A +  +IVSAG+ G I +  E +  N +FA+I
      Sbjct: 117 FPTYLAEEAQYSIIIVSAGKRGMQIELAPEAILS LTNGQFAEI 160

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1061> which encodes the amino acid sequence <SEQ ID 1062>. Analysis of this protein sequence reveals the following:

```

30  Possible site: 30
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2651(Affirmative) < succ>
35  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 114/157 (72%), Positives = 139/157 (87%)

40  Query: 1  MSKKNKIKKTLVDQILDKAKIEHDSLQLDALQGDLPNGIQKQDIFKTLALIGDKTGPIIG 60
      M+KK K+KKTIV+QILDKA I H  L+L+AL+GD P+ +Q  DI+KTLAL GD+TGP+IG
      Sbjct: 1  MAKKTKLKKTLVEQILDKANIAHQGLKLNALGDFPDDLQPSDIYKTLALTGDQTGPLIG 60

45  Query: 61  ILPLTEHLSEKKLAKISGNKKVQMIPQKDLQKITGYIHGANNPIGIRQKHNYPIFIDTIA 120
      I+PLTEHLSEK+LAK+SGNKKV M+PQKDLQK TGYIHGANNP+GIRQKH+YPIFID  A
      Sbjct: 61  IILPLTEHLSEKQLAKVSGNKKVSMVPQKDLQKTTGYIHGANNPVGIRQKHSYPIFIDQTA 120

50  Query: 121 LEKQELIVSAGEIGRSIRINSEVLADFVNAKFADIKE 157
      LEK ++IVSAGE+GRSI+I+S+  LADFV A FAD+K+
      Sbjct: 121 LEKGQIIIVSAGEVGRSIKISSQALADFVGASFADLKK 157

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 328**

A DNA sequence (GBSx0358) was identified in *S.agalactiae* <SEQ ID 1063> which encodes the amino acid sequence <SEQ ID 1064>. Analysis of this protein sequence reveals the following:

-417-

Possible site: 28

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

5                   bacterial cytoplasm --- Certainty=0.4719(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8555> which encodes amino acid sequence <SEQ ID 8556>  
 10 was also identified. This protein belongs to the glycolysis/gluconeogenesis pathway, and such proteins have  
 been experimentally detected as surface-exposed in *Streptococci*. The protein has homology with the  
 following sequences in the GENPEPT database:

>GP:AAD36444 GB:AE001791 phosphoglycerate mutase [Thermotoga maritima]  
 Identities = 65/191 (34%), Positives = 93/191 (48%), Gaps = 13/191 (6%)

15           Query: 5   MKFYLVHRGKTQWNLEGRFQGGANGDSPLLEEAIEELEELGQYLSSIHFDVAVYSSDLGRAR 64  
                   MK YL+RHG+T WN +G +QG   D PL E   E+ +L   L +   DA+YSS L R+  
           Sbjct: 1   MKLYLIRHGGETIWNKGLWQGVTDVPLNERGREQARKLANSLKRV--DAIYSSPLKRSL 57

20           Query: 65   DTVNILNDANSCPKIEHYTPQLREWALGTLEGCKIATMQAIYPRQMTAFYQNPLQFKHDM 124  
                   +T   +   A   KEI   LRE +   G +   YP +   + +P   M  
           Sbjct: 58   ETAEEI--ARRFEKEIIVEEDLRECEISLWNGLTVEEAIREYPVEFKWSSDP---NFGM 112

25           Query: 125 FGAESLYQTTHRVESFLRSLASK----NYDKVLIVGHGANLTASIRSLGQYQYGSLSHYKD 180  
                   G ES+   +RV   + + S+   + V+IV H +L A I +LG   LH  
           Sbjct: 113 EGGLEMRNVQNRVVKAIMKIVSQEKLNGSENVVIVSHSLSLRAFCWILGLPL-YLHRNF 171

30           Query: 181 KLDNASLTIIE 191  
                   KLDNASL+++E  
           Sbjct: 172 KLDNASLSVVE 182

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1065> which encodes the amino acid  
 sequence <SEQ ID 1066>. Analysis of this protein sequence reveals the following:

Possible site: 24

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

40                   bacterial cytoplasm --- Certainty=0.3628(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 127/205 (61%), Positives = 152/205 (73%)

45           Query: 5   MKFYLVHRGKTQWNLEGRFQGGANGDSPLLEEAIEELEELGQYLSSIHFDVAVYSSDLGRAR 64  
                   MK Y VRHGKT WNLEGRFQGA GDSPLLEEA +E+ LG+ L+ + FDAVY+SDL RA  
           Sbjct: 1   MKLYFVRHGKTLWNLEGRFQAGGDSPLLEEAKEIHLGKELAKVAFDAVYTSDLQRAM 60

50           Query: 65   DTVNILNDANSCPKIEHYTPQLREWALGTLEGCKIATMQAIYPRQMTAFYQNPLQFKHDM 124  
                   T I+ DA   ++++T QLREW LG LEG KIATM AIYP+QM AF +N QFK D  
           Sbjct: 61   ATAAIIIDAFDQPKLYHTDQLREWRLGKLEGAKIATMAAIYPQQMLAFRENLAQFKPDQ 120

55           Query: 125 FGAESLYQTTHRVESFLRSLASKNYDKVLIVGHGANLTASIRSLGQYQYGSLSHYKDKLDN 184  
                   F AES+YQTT RV   ++S   K+Y VLI VGHGANLTA+IRSLG++   L K LDN  
           Sbjct: 121 FEAESLYQTTQRVCHLIQSFKDKHYQNVLI VGHGANLTATIRSLGFEPAALLAKGGLDN 180

60           Query: 185 ASLTIIETHDFKDFNCLTWNDKSYL 209  
                   ASLTI+ET D+ ++CL WNDKS+L  
           Sbjct: 181 ASLTILETKDYLTDCLIWNDKSFL 205

SEQ ID 8556 (GBS314) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 4; MW 27.2kDa), in Figure 169 (lane 15-17; MW 41.6kDa) and in Figure 239 (lane 4; MW 41.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 4; MW 52.1kDa).

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 329

A DNA sequence (GBSx0359) was identified in *S.agalactiae* <SEQ ID 1067> which encodes the amino acid sequence <SEQ ID 1068>. Analysis of this protein sequence reveals the following:

```

10 Possible site: 56
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.3014(Affirmative) < succ>
15    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

20 >GP:CAB12562 GB:Z99108 similar to hypothetical proteins [Bacillus subtilis]
    Identities = 69/232 (29%), Positives = 108/232 (45%), Gaps = 9/232 (3%)

    Query: 4  SIVFDVDDTIYDQQAPYRIAVEKCFPDFDMSAINQAYIRFRHYSDIGFPRVMAGEWTTEY 63
    +++FDVDDTI D QA  +A+  F D ++  N  +++  +  +  G+ T +
    Sbjct: 6  TLLFDVDDTILDFQAAEALALRLLFEDQNIPLTNDMKAQYKTINQGLWRAFEEGKMTRDE 65

    Query: 64  FRFWRKETLLEFGYREIDEATGIYFQEIYEHELENITMLDEMRTLDFLKSKNVPMGII 123
    R  L E+GY  EA G  ++ Y  LE  L +  L  +  + I+
    Sbjct: 66  VVNTFRSALLKEYGY----EADGALLEQKYRRFLEBQHQLIDGAFDLISNLQQQFDLYIV 121

    Query: 124 TNGPTEHQLKKVKKLGLYDYVDPKRIVVSQATGFQKPEKEIFNLAAEQF-DMNPSTTLYV 182
    TNG + Q K+++ GL+ +  K + VS+ TGFQKP KE FN  E+  +  TL +
    Sbjct: 122 TNGVSHTYKRLRDSGLFPFF--KDIFVSEDTFQKPMKEYFNYVFERIPQFSAEHTLII 179

    Query: 183 GDSYDNDIMGAFNGGWHSMWFNHRGSLKPGIKPVYDVAIDNFEQLFGAVKV 234
    GDS  DI G  G  + W N  +  P I P Y+  I  E+L+  +  +
    Sbjct: 180 GDSLTAIDIKGQLAGLDTCMWNPDMKPNVPEIIPTYE--IRKLEELYHILNI 229

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1069> which encodes the amino acid sequence <SEQ ID 1070>. Analysis of this protein sequence reveals the following:

```

40 Possible site: 56
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.3216(Affirmative) < succ>
45    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

50 Identities = 276/300 (92%), Positives = 292/300 (97%)

    Query: 1  MITSIVFDVDDTIYDQQAPYRIAVEKCFPDFDMSAINQAYIRFRHYSDIGFPRVMAGEWT 60
    MIT+IVFDVDDTIYDQQAPYRIA+EKCFPDFDMS +NQAYIRFRHYSD+GFPRVMAGEWT
    Sbjct: 1  MITAIVFDVDDTIYDQQAPYRIAMEKCFPDFDMSVMNQAYIRFRHYSDVGFPRVMAGEWT 60

    Query: 61  TEYFRFWRKETLLEFGYREIDEATGIYFQEIYEHELENITMLDEMRTLDFLKSKNVPM 120
    TEYFRFWRKETLLEFGYREIDEA G++FQE+YEHELENITMLDEMRTLDFLKSKNVPM

```

-419-

Sbjct: 61 TEYFRFWRKETLLEFGYREIDEAAGVHFQEVYEHLENITMLDEMRTLDLFLKSKNVPM 120

Query: 121 GIITNGPTEHQLKKVKKLGLYDYDVPKRIVVSQATGFQKPEKEIFNLAAEQFDMNPSTTL 180  
GIITNGPTEHQLKKV+KLGLYDY+D KRVIVVSQATGFQKPEKEIFNLAAEQFDMNP TTL

5 Sbjct: 121 GIITNGPTEHQLKKVRKLGLYDYIDAKRVIVVSQATGFQKPEKEIFNLAAEQFDMNPQTTL 180

Query: 181 YVGDSYDNDIMGAFNGGWHSMWFNHRGRSLKPGIKPVYDVAIDNFEQLFGAVKVLFDLPD 240  
YVGDSYDNDIMGAFNGGWHSMWFNHRGR LKPG KPVYDVAIDNFEQLFGAVKVLFDLPD

10 Sbjct: 181 YVGDSYDNDIMGAFNGGWHSMWFNHRGRQLKPGTKPVYDVAIDNFEQLFGAVKVLFDLPD 240

Query: 241 NKFIFDINDKSNPVLEMGLNNGLMMAAERLLESNMSVDKVILLRLTAKQEKVLRMKYAR 300  
NKFIFD+NDK NP+L+MG+NNGLMMAAERLLESNMS+DKVILLRLT +QEKVLR+KYAR

Sbjct: 241 NKFIFDVNDKKNPILQMGINNGLMMAAERLLESNMSIDKVILLRLTKQEKVLRMLKYAR 300

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 330

A DNA sequence (GBSx0360) was identified in *S.agalactiae* <SEQ ID 1071> which encodes the amino acid sequence <SEQ ID 1072>. Analysis of this protein sequence reveals the following:

20 Possible site: 34  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2451(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9727> which encodes amino acid sequence <SEQ ID 9728> was also identified.

30 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB11858 GB:Z99104 lysyl-tRNA synthetase [Bacillus subtilis]  
Identities = 318/490 (64%), Positives = 390/490 (78%), Gaps = 1/490 (0%)

35 Query: 44 EELNDQQIVRREKMAALTEQGIDPFGKRFERTATSGQLNEKYADKSKEDLHDIEETATIA 103  
EELNDQ VRR+KM L + GIDPFG RFERT S ++ Y D +KE+L + TIA  
Sbjct: 9 EELNDQLQVRRDKMNQLRDNGIDPFGARFERTHQSQEVISAYQDLTKEELEEKAIEVTIA 68

Query: 104 GRLMTKRKGKGVGFAGHIQDREGQIQIYVRKDSVGEENYEIFKKADLGDFLGVEGQVMRTD 163  
GR+MTKRKGKG GFAH+QD EGQIQIYVRKDSVG++ YEIFK +DLGD +GV G+V +T+

40 Sbjct: 69 GRMMTKRKGKAGFAHLQDLEQIQIYVRKDSVGDDQYEIFKSSDLGLIGVTGKVFKN 128

Query: 164 MGELSIKATHITHLSKALRPLPEKFHGLTDIETIYKRHLDLISNRDSFDRFVTRSKIIS 223  
+GELS+KAT L+KALRPLP+K+HGL D+E YR+R+LDLI N DS F+TRSKII

45 Sbjct: 129 VGELSVKATSFELLTKALRPLPDKYHGLKDVEQRYRQRYLDLIVNPDSKHTFITRSKIIQ 188

Query: 224 EIRRFMDNSNGFLEVETPVLHNEAGGASARPFITHNAQDIDMVLRIATELHLKRLIVGGM 283  
+RR++D +G+LEVETP +H+ GGASARPFITHNA DI + +RIA ELHLKRLIVGG+

Sbjct: 189 AMRRYLDHGHYLEVETPTMHSIPGGASARPFITHNALDIPLYMRIAELHLKRLIVGGL 248

50 Query: 284 ERVYEIGRIFRNEGMDATHNPEFTSIEAYQAYADYQDIMDLTEGIIQHVTKTIVKGDGPIN 343  
E+VYEIGR+FRNEG+ HNPEFT IE Y+AYADY+DIM LTE ++ H+ + V G I

Sbjct: 249 EKVYEIGRVFRNEGSTRHNPEFTMIELYEAYADYKDIMSLENLVAHIAQEVGLGTTTIQ 308

Query: 344 YQCTEIKINEPFRVHMVDVAVKEITGIDFWKEMTLEEAQALAEKNVPLEKHFTTVGHII 403  
Y +I + +KR+HMVDVAVKE TG+DFW+E+T+E+A+ A+E V + K TVGHII

55 Sbjct: 309 YGEEQIDLKPEWKRIHMVDVAVKEATGVDFWEEVTVQAREYAKEHEVEI-KDSMTVGHII 367

Query: 404 NAFEEFVEDTLIQPTFVFGHPVEVSPLAKKNDTPRFTDRFELFIMTKEYANAFTELND 463  
N FFE+ +E+TLIQPTF++GHPVE+SPLAKN DPRFTDRFELFI+ +E+ANAFTELND

60 Sbjct: 368 NEFFEQKIEETLIQPTFIYGHVPEISPLAKKNPEDPRFTDRFELFIVGREHANAFTELND 427

-420-

Query: 464 PIDQLSRFEAQASAKELGDDEATGVDYDYVEALEYGPPTGGLGIGIDRLCMLLTDTTTI 523  
 PIDQ RFEAQ +E G+DEA +D D+VEALEYGPPTGGLGIGIDRL MLLT+ +I  
 Sbjct: 428 PIDQRERFEAQLKEREAGNDEAHLMDDEFVEALEYGPPTGGLGIGIDRLVMLLTNPAPSI 487

Query: 524 RDVLLFPTMK 533  
 RDVLLFP M+  
 Sbjct: 488 RDVLLFPQMR 497

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1073> which encodes the amino acid sequence <SEQ ID 1074>. Analysis of this protein sequence reveals the following:

Possible site: 45  
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4694(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below:

Identities = 439/500 (87%), Positives = 474/500 (94%)

- Query: 34 LEEIMSNQHIEELNDQQIVRREKMAALTEQGIDPFGKRFERTATSGQLNEKYADKSKEDL 93  
 LEE MSNQHIEELNDQQIVRREKM AL EQGIDPFGKRF+RTA S +L EKYADK+KE+L  
 Sbjct: 1 LEENMSNQHIEELNDQQIVRREKMTALAEQGIDPFGKRFDR+TANS AELKEKYADKTKEEL 60
- Query: 94 HDIEETATIAGRLMTKRGKKGKVGFAHIQDREGQIQIYVRKDSVGEENYEI+FKKADLGDFL 153  
 H++ ETA +AGRLMTKRGKKGKVGFAH+QDREGQIQ+YVRKDSVGE+NYE+I+FKKADLGDF+  
 Sbjct: 61 HELNETAIVAGRLMTKRGKKGKVGFAHLQDREGQIQIYVRKDSVGEDNYE+I+FKKADLGDFI 120
- Query: 154 GVEGQVMRTDMGELSIKATHITHLSKALRPLPEKFHGLTDIETIYKRHLDLISNRD SFD 213  
 GVEG+VMRTDMGELSIKAT +THLSK+LRPLPEKFHGLTDIETIYKRHLDLISNR+SFD  
 Sbjct: 121 GVEGEVMRTDMGELSIKATKLTHLSKSLRPLPEKFHGLTDIETIYKRHLDLISNR SFD 180
- Query: 214 RFVTRSKIISEIRRFMD SNGFLEVETPVLHNEAGGASARPFITHHNAQDIDMVLRIATEL 273  
 RFVTRSK+ISEIRR++D FLEVETPVLHNEAGGA+ARPF+THHNAQ+IDMVLRIATEL  
 Sbjct: 181 RFVTRSKMISEIRRYLDGLDFLEVETPVLHNEAGGAAARPFVTHHNAQNIDMVLRIATEL 240
- Query: 274 HLKRLIVGGMERVYEIGRIFRNEGMDATHNPEFTSIEAYQAYADYQDIMDLTEGIIQHVT 333  
 HLKRLIVGGMERVYEIGRIFRNEGMDATHNPEFTSIE YQAYADY DIM+LTEGIIQH  
 Sbjct: 241 HLKRLIVGGMERVYEIGRIFRNEGMDATHNPEFTSIEVYQAYADYLDIMNLTGIIQHAA 300
- Query: 334 KTVKGDGPINYGQTEIKINEPFRVHMVDAVKEITGIDFWKEMTLEEAQALAEKQNVPLE 393  
 K V+GDGPI+YQGTEI+INEPFRVHMVDA+KE+TG DFW EMT+EEA ALA+EK VPLE  
 Sbjct: 301 KAVRGDGPIDYQGTEIRINEPFRVHMVD AKEVTGADFWPEMTVEEAIALAKEKQNVPLE 360
- Query: 394 KHFTTVGHIINAFFEEFVEDTLIQPTFVFGHPVEVSPLAKKNDTDRFTDRFELFIMTKE 453  
 KH F +VGHIINAFFEEFVE+TL+QPTFVFGHPVEVSPLAKK N D RFTDRFELFIMTKE  
 Sbjct: 361 KH F ISVGHIINAFFEEFVEETLVQPTFVFGHPVEVSPLAKKNPEDTRFTDRFELFIMTKE 420
- Query: 454 YANAFTELNDPIDQLSRFEAQASAKELGDDEATGVDYDYVEALEYGPPTGGLGIGIDRL 513  
 YANAFTELNDPIDQLSRFEAQ A KEGDDEATG+DYD+VEALEYGPPTGGLGIGIDRL  
 Sbjct: 421 YANAFTELNDPIDQLSRFEAQAKELGDDEATGIDYDFVEALEYGPPTGGLGIGIDRL 480
- Query: 514 CMLLTDTTTTIRDVLLFPTMK 533  
 CMLLT+TTTIRDVLLFPTMK  
 Sbjct: 481 CMLLTNTTTTIRDVLLFPTMK 500

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-421-

**Example 331**

A DNA sequence (GBSx0361) was identified in *S.agalactiae* <SEQ ID 1075> which encodes the amino acid sequence <SEQ ID 1076>. This protein is predicted to be 6,7-dimethyl-8-ribityllumazine synthase (ribH). Analysis of this protein sequence reveals the following:

```

5   Possible site: 34
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.1042(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15  >GP:CAB14257 GB:Z99116 riboflavin synthase (beta subunit) [Bacillus subtilis]
    Identities = 103/151 (68%), Positives = 120/151 (79%)

    Query: 1   MTIEGQLVANEMKIGIVVSRFNEIITSKLLSGAVDGLLRHGVSEEDIDIVWVPGAFEIP 60
              M II+G LV  +KIGIVV RFN+ ITSKLLSGA D LLRHGV  DID+ WVPGAFEIP
    Sbjct: 1   MNIIQGNLVGTGLKIGIVVGRFNFITSKLLSGAEDALLRHGVDTNDIDVAWVPGAFEIP 60

20  Query: 61   YMARKMALYKDYDAIICLVVIGKSTHDYDYVCNEVTKGIGHLSQSDIPHIFGVLTTDN 120
              + A+KMA K YDAII LG VI+G+T HYDYVCNE KGI  + + +P IFG++TT+N
    Sbjct: 61   FAAKKMAETKKYDAIITLGTVIRGATTHYDYVCNEAAKGIAQAANTTGVVPVIFGIVTTEN 120

25  Query: 121  IEQAIERAGTKAGNKG YDCALSAIEMVNLDK 151
              IEQAIERAGTKAGNKG DCA+SAIEM NL++
    Sbjct: 121  IEQAIERAGTKAGNKG VDCAVSAIEMANLNR 151

```

No corresponding DNA sequence was identified in *S.pyogenes*.

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 332**

A DNA sequence (GBSx0362) was identified in *S.agalactiae* <SEQ ID 1077> which encodes the amino acid sequence <SEQ ID 1078>. This protein is predicted to be GTP cyclohydrolase ii (ribA/B). Analysis of this protein sequence reveals the following:

```

35  Possible site: 20
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
40          bacterial cytoplasm --- Certainty=0.1918(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

45 A related GBS nucleic acid sequence <SEQ ID 9725> which encodes amino acid sequence <SEQ ID 9726> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

50  >GP:AAA86524 GB:U27202 GTP cyclohydrase II/
      3,4-dihydroxy-2-butanone-4-phosphate synthase
      [Actinobacillus pleuropneumoniae]
    Identities = 230/395 (58%), Positives = 307/395 (77%)

    Query: 19   FSPIKKLLQDIKSGKMVVLMDDENRENEGD LICAEMVTKESINFMAKFGKGLICLPLSN 78
              FS ++ ++ I+ GK++++ DDE+RENEGD ICAAE T E+INFMA +GKGLIC P+S
    Sbjct: 6   FSKVEDAIEAIRQGKIILVTDDRENEGDFICAAEFATPENINFMATYKGGLICTPIST 65

```

-422-

Query: 79 YYAEKLELAQMASHNTDNHETAFTISIDHLSTSTGISAE DRALTAKMVANDSSKAKDFRR 138  
 A+KL M + N DNHETAFT+S+DH+ T TGISA +R++TA + +D++KA DFRR  
 5 Sbjct: 66 EIAKKLNFMVAVNQDNHETAFTVSVDHIDTGTGISAFERSITAMKIVDDNAKATDFRR 125

Query: 139 PGHLPFLLAKEGGVLRNGHTEATVDLCRLAGLKECGLCCEIMAEDGSMMRKDELLAFAQ 198  
 PGH+FPL+AKEGGVL RNGTHTEATVDL RLAGLK GLCCEIMA+DG+MM +L FA  
 Sbjct: 126 PGHMFPLIAKEGGVLRNGHTEATVDLARLAGLKHAGLCCEIMADDGTMMPDLQKFAV 185

10 Query: 199 KHDLAIAITIKQLQDYRRQE EGGVREIEIQLPTQFGHFTAYGYSEVVANKEHVALVKGDI 258  
 +H++ TI+QLQ+YRR+ + V + +++PT++G F A+ + EV++ KEHVALVKGD+  
 Sbjct: 186 EHNMPFITIQQLQEYRRKHDLSLVKQISVVKMPTKYGEFMAHSFVEVISGKEHVALVKGDL 245

15 Query: 259 SSGEDVLCRLHSECLTGDV FHSRLCDCGEQLANALQQIEA EGRGVLLYMRQEGRGIGLIN 318  
 + GE VL R+HSECLTGD F S RCDG+Q A A+ QIE EGRGV+LY+RQEGRGIGLIN  
 Sbjct: 246 TDGEQVLARIHSECLTGD AFGSQRCDGQQAAMTQIEQEGRGVILYLRQEGRGIGLIN 305

Query: 319 KLKAYHLQE EGLDTLEANLALGFEGDERDYGVS AQLLKDLGINSINLLTNNPDKIQOLEA 378  
 KL+AY LQ++G+DT+EAN+ALGF+ DER+Y + AQ+ + LG+ SI LLTNNP KI+ L+  
 20 Sbjct: 306 KLRAYELQDKGMDTVEANVALGFKEDEREYYIGAQMFOQLGVKSIRLLTNNPAKIEGLKE 365

Query: 379 EGICVKNRVPLQVAVTAYDLNLYLTKKEKMGHLLD 413  
 +G+ + R P+ V D++YLK K+ KMGH+ +  
 25 Sbjct: 366 QGLNIVAREPIIVEPNKNDIDYLVKQIKMGHMFN 400

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 333

30 A DNA sequence (GBSx0363) was identified in *S.agalactiae* <SEQ ID 1079> which encodes the amino acid sequence <SEQ ID 1080>. This protein is predicted to be riboflavin synthase alpha chain (ribE). Analysis of this protein sequence reveals the following:

Possible site: 59  
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3517(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9723> which encodes amino acid sequence <SEQ ID 9724> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

45 >GP: BAB05274 GB: AP001512 riboflavin synthase alpha subunit [Bacillus halodurans]  
 Identities = 98/216 (45%), Positives = 147/216 (67%), Gaps = 2/216 (0%)

Query: 1 MFTGIIIEEMQVSRIRNGIKSQQLSIDAPKLVPLLRKGDVAVNGVCLTVLDKSETAFIA 60  
 MFTGIIIE++G + I+ ++ ++I + K+V ++ GDS+AVNGVCLTV ++T F  
 Sbjct: 1 MFTGIIIEDVGTIDAIQQTGEAIVMTTTSKKIVSDVQLGDSIAVNGVCLTVTSFTDTQFTV 60

50 Query: 61 DVMPESMRMTSLAALRLHSAKVNLELALRSDSRLGGHFVLGHVDGVGKIEKIQKDDIAVRF 120  
 D+MPE++ TSL L S+VNLE A+ ++ R GGH V GHVDG+G I K ++ D AV +  
 Sbjct: 61 DLMPEVTRATSLRLLSKGSRVNLERAMVANGRFGGHIVSGHVDGIGTIRKKERKDNNAVY 120

55 Query: 121 SIDAPPSIMSYIIIEKGSVALDGLSLTVVSFTEHSFEVSVIPHTMAQTNLSLKVKVDLLNI 180  
 +I+ S+ Y+I KGSVA+DG SLT+ ++ +F +S+IPHTM +T + LKK GD++NI  
 Sbjct: 121 TIEVSSSLRRYMIHKGSVAVDGTSLSLTIFDVSDKTFITISIIPHTMEETIIGLKKAGDIVNI 180



-423-

Query: 181 EVDVLGKYAEKFLAPTNRNTHTSSVMDWSFLSENGY 216  
 E D++GKY E+F+ N + +FL+E+GY  
 Sbjct: 181 ECDLIGKYIEQFVQGGKPVNEGG--LTKAFLTEHGY 214

5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 334

10 A DNA sequence (GBSx0364) was identified in *S.agalactiae* <SEQ ID 1081> which encodes the amino acid sequence <SEQ ID 1082>. This protein is predicted to be riboflavin-specific deaminase (ribD). Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.01 Transmembrane 307 - 323 ( 307 - 323)  
 15 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1404(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>  
 20

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA86522 GB:U27202 riboflavin-specific deaminase [Actinobacillus  
 pleuropneumoniae]  
 Identities = 182/353 (51%), Positives = 259/353 (72%)  
 25 Query: 6 DYMALALKEAEKGMGFVAPNPLVGAVIVKDDRIISKGYHKRFGDLHAERQAIKNADEDIS 65  
 DYM A+ A++G+G+ PNPLVG VIVK+ I+++GYH++ G HAER A+ + ED+S  
 Sbjct: 51 DYMRAIALAKQGLGWTNPPLVGCIVKNGEIVAEGYHEKIGGWHAERNAVLHCKEDLS 110  
 30 Query: 66 GSTLYVTLEPCCHVGKQPPCTEALIKSGIKKVVGSLDPNPLVSGKGIALLRKEGLNVEV 125  
 G+T YVTLEPCCH G+ PPC++ LI+ GIKKV +GS DPNPLV+G+G LR+ G+ V  
 Sbjct: 111 GATAYVTLEPCCHHGRTPPCSDLLIERGIKKVFIGSSDPNPLVAGRGANQLRQAGVEVVE 170  
 35 Query: 126 GILREECDALNERFIFHMTYKQPFVYLKYAMTLDGKIATKTGDSKWISNEHSRQSVQKLR 185  
 G+L+EECDALN F ++ K+P+V +KYAMT DGKIAT +G+SKWI+ E +R VQ+ R  
 Sbjct: 171 GLLKEECDALNPIFFHYIQTFRPYVLMKYAMTADGKIATGSGESKWITGESARARVQQTR 230  
 40 Query: 186 QKCSAIMVGINTVLADNPRLTCRIPKGEALVRIVCDSQLKIPLDSYLVKSAKTIPTWIAT 245  
 + SAIMVG++TVLADNP L R+P + VRIVCDSQL+ PLD LV++AK T IAT  
 Sbjct: 231 HQYS AIMVGVDTVLADNPMLNSRMPNAKQPVIVCDSQLRTPLDLQVQTAKEYRTVIAT 290  
 Query: 246 CSDNLAQQQTLKEMGCRLIKVPKDGKLDLKVMTILQGEGIDSLLEGGSSLFHFSALKA 305  
 SD+L + + + +G ++ ++ ++DL+ L+ LG+ IDSL+EGGSSL+FSAL++  
 Sbjct: 291 VSDDLQKIEQFRPLGVDVLVCKARNKRVLDLQDLQKLGEMQIDSLLEGGSSLNFSALSALES 350  
 45 Query: 306 GIVNRLIVFIAPKIIGGLKAKTAISGEGLDWNQAFRVKDIELSRMDSVDVIE 358  
 GIVNR+ +IAPK++GG +AKT I GEG+ ++QA ++K + D++++  
 Sbjct: 351 GIVNRVHCYIAPKLVGKQAKTPIGGEGIQIDQAVKLKLKSTELIGEDILLD 403

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1083> which encodes the amino acid sequence <SEQ ID 1084>. Analysis of this protein sequence reveals the following:

Possible site: 25  
 >>> Seems to have no N-terminal signal sequence  
 55 INTEGRAL Likelihood = -1.17 Transmembrane 88 - 104 ( 88 - 105)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1468(Affirmative) < succ>

-424-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP:CAB11794 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]  
 Identities = 71/161 (44%), Positives = 109/161 (67%)

Query: 13 LEEQTYFMQEALKEAEKSLQKAEIPIGCVIVKDGEEIIGRGNAREESNQAIMHAEMMAIN 72  
 + + +M+EA+KEA+K+ +K E+PIG V+V +GEII R HN RE ++I HAEM+ I+  
 10 Sbjct: 1 MTQDELYMKEAIKEAKKAEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVID 60

Query: 73 EANAHEGNWRLDITTLFVTIEPCVMCSGAIGLARIPHVYIGASNQKFGGVDSLYQILTDE 132  
 EA G WRL TL+VT+EPC MC+GA+ L+R+ V++GA + K G +L +L +E  
 15 Sbjct: 61 EACKALGTWRLEGATLYVTLEPCPMCAVAVLSRVEKVVFGAFDPKGGCSGTLMNLLQEE 120

Query: 133 RLNHRVQVERGLLAADCANIMQTFFRQGRERKKIAKHLIKE 173  
 R NH+ +V G+L +C ++ FFR+ R++KK A+ + E  
 Sbjct: 121 RFNHQAEVVSGLVEECGMLSAFFRELKKKKAARKNLSE 161

20 An alignment of the GAS and GBS proteins is shown below:

Identities = 48/146 (32%), Positives = 71/146 (47%), Gaps = 21/146 (14%)

Query: 7 YMALALKEAEKGMGFVAPNPLVGAVIVKDDRIISKGYHKRFGD---LHAERQAIKNADE 62  
 +M ALKEAEK + A P +G VIVKD II +G++ R +HAE AI A+  
 25 Sbjct: 19 FMQEALKEAEKSLQ-KAEIP-IGCVIVKDGEEIIGRGNAREESNQAIMHAEMMAINEANA 76

Query: 63 D-----ISGSTLYVTLEPCCHVGKQPPCTEALIKSGIKKVVVGS LDPNPLVSGKGIALLR 117  
 + +TL+VT+EPC C+ A+ + I V+ G+ + +L  
 30 Sbjct: 77 HEGNWRLDITTLFVTIEPCV-----MCSGAIGLARIPHVYIGASNQKFGGVDSLYQILT 130

Query: 118 KEGLN----VEVGILREECDALNERF 139  
 E LN VE G+L +C + + F  
 Sbjct: 131 DERLNHRVQVERGLLAADCANIMQTF 156

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 335

A DNA sequence (GBSx0365) was identified in *S.agalactiae* <SEQ ID 1085> which encodes the amino acid sequence <SEQ ID 1086>. This protein is predicted to be Nramp metal ion transporter. Analysis of this  
 40 protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

45 INTEGRAL Likelihood = -11.89 Transmembrane 169 - 185 ( 160 - 191)  
 INTEGRAL Likelihood = -11.09 Transmembrane 140 - 156 ( 128 - 165)  
 INTEGRAL Likelihood = -6.85 Transmembrane 359 - 375 ( 354 - 379)  
 INTEGRAL Likelihood = -6.48 Transmembrane 269 - 285 ( 263 - 287)  
 INTEGRAL Likelihood = -6.16 Transmembrane 426 - 442 ( 423 - 445)  
 INTEGRAL Likelihood = -5.57 Transmembrane 62 - 78 ( 58 - 80)  
 INTEGRAL Likelihood = -4.94 Transmembrane 107 - 123 ( 103 - 127)  
 50 INTEGRAL Likelihood = -4.46 Transmembrane 391 - 407 ( 389 - 408)  
 INTEGRAL Likelihood = -4.35 Transmembrane 310 - 326 ( 307 - 328)

----- Final Results -----

55 bacterial membrane --- Certainty=0.5755(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF83825 GB:AE003939 manganese transport protein [Xylella

-425-

fastidiosa]

Identities = 192/436 (44%), Positives = 274/436 (62%), Gaps = 14/436 (3%)

5 Query: 10 SLSEVNQSVEVPHNSSFWNTLRAFLGPGALVAVGYMDPGNWITSVIGGATYRYLLLFVVL 69  
 SL E++ SV V + L AFLGPG +V+VGYMDPGNW T + GG+ + Y+LL V+L  
 Sbjct: 39 SLGEMHASVAVSRRGHGFRLLAFLGPGYMVSVGYMDPGNWATGLAGGSRFYGYMLLSVIL 98

10 Query: 70 VSSLAMQLQQMAGKLGIVTRQDLAQATASRLPKPLRYLLFIIELALIAITDLAEVIGSA 129  
 +S++MA+ LQ +A +LGI + DLAQA +R + L+++ ELA+IA DLAEVIG+A  
 Sbjct: 99 LSNVMAIVLQALAARLGIASDMDLAQACRARSRGTTIALWVVCELAIACDLAEVIGTA 158

15 Query: 130 IALHLLFGWPILLSIMITILDVFLLLMLKLGQKIEAFVSVLILITILIIIFTYLVVLSQP 189  
 IAL+LL G P++ ++IT +DV L+LLLM G + +EAFV L+L I F +VL+ P  
 Sbjct: 159 IALNLLGVPPIIWGVVITAVDVVLVLLMHRGFRALEAFVIALLLVIFGCFVQIVLAAP 218

20 Query: 190 DLDAMFKGFLPHHELFNISHEGKNSPLTLALGIIGATVMPHNLYLHSSLSQTRRDYHKN 249  
 L + GF+P ++ L LA+GI+GATVMPHNLYLHSS+ QTR  
 Sbjct: 219 PLQEVLLGGFVPRWQVV-----ADPQALYLAIGIVGATVMPHNLYLHSSIVQTRAYP-RTP 272

25 Query: 250 SSIKKAVERFMTLDSNIQLSLAFVNVNLLLVLGASLFYG-HANDISAFSOMYLALSDKTIT 308  
 + A+R+ DS + L LA +N+ +L+L A++F+ H D+ Q Y L+  
 Sbjct: 273 VGRRSALRWAVADSTLALMLALFINASILILAAAVFHAQHHDVEEIEQAYQLLAPVLGV 332

30 Query: 309 GAVASSFLSTLFAVALLASGQNSTITGTLTGQIVMEGFLHFKLPQWLIRLCTRLLTLPI 368  
 G A TLFA ALLASG NST+T TL GQIVMEGFL +L WL R+ TR L ++P+  
 Sbjct: 333 GVAA-----TLFATALLASGINSTVTATLAGQIVMEGFLRLRLRPWLRRVLTGLAIVPV 387

35 Query: 369 FVIALLVGGEENTLDQLIVYSQVFLSLALPFSIFPLIYFTSQKSIMGEHANAKWNTYLAY 428  
 V+ L G E +L++ SQV LS+ LPF++ PL+ + + +MG +W +A+  
 Sbjct: 388 IVVVALYG--EQGTGRLLLLSQVILSMQLPFAVIPLLRVADRKMVGALVAPRWLMVVAW 445

Query: 429 LVAILTLNLKLIMD 444  
 L+A ++ +LN+KL+ D  
 Sbjct: 446 LIAGVIVVLNVKLLGD 461

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 336

40 A DNA sequence (GBSx0366) was identified in *S.agalactiae* <SEQ ID 1087> which encodes the amino acid sequence <SEQ ID 1088>. Analysis of this protein sequence reveals the following:

Possible site: 54

&gt;&gt;&gt; Seems to have a cleavable N-term signal seq.

45 INTEGRAL Likelihood = -14.12 Transmembrane 113 - 129 ( 98 - 132)  
 INTEGRAL Likelihood = -12.15 Transmembrane 228 - 244 ( 220 - 249)  
 INTEGRAL Likelihood = -10.83 Transmembrane 175 - 191 ( 167 - 195)  
 INTEGRAL Likelihood = -5.04 Transmembrane 57 - 73 ( 55 - 75)  
 INTEGRAL Likelihood = -3.93 Transmembrane 146 - 162 ( 142 - 166)  
 50 INTEGRAL Likelihood = -1.38 Transmembrane 199 - 215 ( 199 - 215)  
 INTEGRAL Likelihood = -0.32 Transmembrane 82 - 98 ( 82 - 98)

----- Final Results -----

55 bacterial membrane --- Certainty=0.6647(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF11325 GB:AE002018 hypothetical protein [Deinococcus radiodurans]  
 Identities = 63/215 (29%), Positives = 108/215 (49%), Gaps = 13/215 (6%)

60 Query: 11 LLLVFILTIIVNYLSATGFLTGNSQKSLSDRYQTLTPAPLAFSIWSVIYL-LTFLVILR 69

-426-

LL +LT++VNYLS L GNS +SDR TPA L F++W I+L L ++  
 Sbjct: 10 LLAATVLTlLVNylSNALPLFGNSAEVSDRLPNAFTPAGLTFTVWGPIFLGLLVFAVYQ 69  
 Query: 70 AIFSksQSYQDNFASIFPYFLGllLVNNIWTVFFTSNLIGLSTIIIFAYCILLV-IIIKI 128  
 A+ ++ + D +P+ LG LL N W + F S IGLS +I+ A +LV + + +  
 Sbjct: 70 ALPAQRGARLDRL--FWPFLLGnLL-NVAWLLAFQSLNIGLSVIMLALLAVLVRLYLSV 126  
 Query: 129 LS---KNKSKLLLRITFGIHAGWLLVASLVNLAVYLVKI---DFNYPLPKVYIAIIALI 181  
 S + + L++ ++ W+ VA++ N+ +LV F V+ A++ ++  
 Sbjct: 127 RSLPPQGAERWTLQLPVSLYLAWISVATIANITAFVLSAGVTQSFLGIAGPVWSALLLV 186  
 Query: 182 FITVLSLYLARVLQNAVILILSVFWAWLMVFAHLE 216  
 + +L R A+ + + WA+ V+ A E  
 Sbjct: 187 AAAIGVFFLWRFRDYAFAAV-LLWAFYGVYVARPE 220

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 337

20 A DNA sequence (GBSx0367) was identified in *S.agalactiae* <SEQ ID 1089> which encodes the amino acid sequence <SEQ ID 1090>. Analysis of this protein sequence reveals the following:

Possible site: 36  
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.3401(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC65352 GB:AE001215 T. pallidum predicted coding region  
 TP0352 [Treponema pallidum]  
 Identities = 28/64 (43%), Positives = 41/64 (63%)

35 Query: 3 EFTFEIVEKLLVLSENEKGWTKELNRSVFNGAPAKFDLRTWSPDHTKMGKGITLSNEEFK 62  
           +F +E+ LS + GW+ EL +S+NG P K+D+R WSPD +KMGKG+TL+ E  
 Sbjct: 12 DFHYEVTRNWGTLSTSGNGWSLELKSISWNGRPEKYDIRAWSPDKSKMGKGVTLTRAEIV 71

40 Query: 63 VILD 66  
           + D  
 Sbjct: 72 ALRD 75

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1091> which encodes the amino acid sequence <SEQ ID 1092>. Analysis of this protein sequence reveals the following:

45 Possible site: 36  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 50           bacterial cytoplasm --- Certainty=0.4021(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

55 Identities = 59/70 (84%), Positives = 64/70 (91%)  
 Query: 1 MSEFTFEIVEKLLVLSENEKGWTKELNRSVFNGAPAKFDLRTWSPDHTKMGKGITLSNEE 60  
           M+EFTF I E LL LSEN+KGWTKELNRSVFNGA AK+D+RTWSPDHTKMGKGITL+NEE  
 Sbjct: 1 MAEFTFNIEHLLTLSENDKGWTKELNRSVFNGAEAKWDIRTWSPDHTKMGKGITLTNEE 60

-427-

Query: 61 FKVILDAFRK 70  
 FK ILDAFRK  
 Sbjct: 61 FKTILDAFRK 70

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 338

A DNA sequence (GBSx0368) was identified in *S.agalactiae* <SEQ ID 1093> which encodes the amino acid sequence <SEQ ID 1094>. Analysis of this protein sequence reveals the following:

Possible site: 61  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -2.66 Transmembrane 92 - 108 ( 92 - 110)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2062(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14676 GB:Z99117 similar to protease [Bacillus subtilis]  
 Identities = 201/407 (49%), Positives = 277/407 (67%), Gaps = 2/407 (0%)

Query: 4 VKKRPEVLSPAGTLEKLKVAIDYGADAVFVGGQAYGLRSRAGNFSMEELQEGINYAHARD 63  
 + K+PE+L+PAG LEKLK+A+ YGADAVF+GGQ YGLRS A NF++EE+ EG+ +A  
 Sbjct: 18 ITKKPELLAPAGNLEKLKIAVHYGADAVFIGGQEQYGLRSNADNFTIEEIAEGVEFAKKYG 77

Query: 64 AKVYVAAVMVTHEGNELGAGPWFRELDMGLDAVIVSDPALIVICATEAPGLEIHLSTQA 123  
 AK+YV N+ H N G + + L D + +IV+DP +I C AP +E+HLSTQ  
 Sbjct: 78 AKIYVITNIFAHNENMDGLELYLKALGDANVAGIIVADPLIETCRRVAPNVEVHLSTQQ 137

Query: 124 SSTNYETFEFWKEMGLTRVVLAREVTMAELAEIRKRTDVEIEAFVHGAMCISYSGRCVLS 183  
 S +N++ +FWKE GL RVVLARE + E+ E++++ D+EIE+F+HGAMCI+YSGRCVLS  
 Sbjct: 138 SLSNWKAVQFWKEGLDRVVLARETSALEIREMKEKVDIEIESFIHGAMCIAYSGRVLS 197

Query: 184 NHMSHRDANRGGCSQSCRWKYDLYDMPFGQERQSLKGEIPEPFSMSAVDMCMIEHIPDMI 243  
 NHM+ RD+NRGGC QSCRW YDLY G +L GE PF+MS D+ +IE IP MI  
 Sbjct: 198 NHMTARDSNRGGCCQSCRWDYDLYQTD-GANAVALYGEEDAPFAMSPKDLKLIESIPKMI 256

Query: 244 ENGVDLSLKIEGRMKSIIHYVSTVTNCYKAAYDAYMESPEAFEAIKEDLIDELWKVAQRELA 303  
 E G+DSLKIEGRMKSIIHYV+TV + Y+ +DAY PE F I+++ ++EL K A R+ A  
 Sbjct: 257 EMGIDLSLKIEGRMKSIIHYVATVVSVRKVIDAYCADPENF-VIQEWLEELDCKANRDTA 315

Query: 304 TGFYHTPTENEQLFGARRKIPQYKFVGEVVSFDNAKMEATIRQRNVIMEGDRVEFYGP 363  
 T F+ TP EQ+FG K Y FVG V+++D T++QRN +GD VEF+GP  
 Sbjct: 316 TAFEGTTPGYEEQMFGEHAKKTTYDFVGLVLNYDEDTQMVTIQQRNFFKKGDEVEFFGP 375

Query: 364 FRHFECFIDGLRDAEGNKIDRAPNPMELLTITLNPVKKGDMIRACK 410  
 +F I+ + D +GN++D A +P++++ L + +M+R K  
 Sbjct: 376 IENFTHTIETIWDGDNELDAARHPLQIVKFKLDKKIYPSNMMRKKG 422

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1095> which encodes the amino acid sequence <SEQ ID 1096>. Analysis of this protein sequence reveals the following:

Possible site: 61  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -2.66 Transmembrane 92 - 108 ( 92 - 110)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2062(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-428-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the databases:

&gt;GP:BAB04993 GB:AP001511 protease [Bacillus halodurans]

Identities = 201/403 (49%), Positives = 280/403 (68%), Gaps = 4/403 (0%)

Query: 6 KRPEVLSPAGTLEKLKVAIDYGADAVFVGQAYGLRSRAGNFSMEELQEGIDYAHARGAK 65

K+PE+L+PAG+LEKLKVAI YGADAV++GGQ +GLRS A NFS+EE++EG+++A+ GAK

Sbjct: 17 KKPELLAPAGSLEKLKVAIHGADAVYIGGQEFGLRSNADNFSIEMREGVEFANKYGAK 76

Query: 66 VYVAANMVTHEGNEIGAGEWFRQLRDMGLDAVIVSDPALIVICSTEAPGLEIHLSTQASS 125

VYV N+ H N G E+ L+++G+ +IV+DP +I C AP +E+HLSTQ S

Sbjct: 77 VYVTNNIYAHNENMDGLEEYLSALQEVGVGTGIIVADPLIETCKRVAPKVEVHLSTQQSL 136

Query: 126 TNYETFEFWKAMGLTRVVLAREVNMAELAEIRKRTDVEIEAFVHGAMCISYSGRCVLSNH 185

+N+ +FWK GL RVVLAREV + E+ E++K D+EIE FVHGAMCISYSGRCVLSNH

Sbjct: 137 SNWLAVKFWKEGLHRVVLAREVGLLEEMLEMKHVDIEIETFVHGAMCISYSGRCVLSNH 196

Query: 186 MSHRDANRGGCSQSCRWKYDLYDMPFGGE-RRSLKGEIPEDYSMSVDMCMIDHIPDLIE 244

M+ RD+NRGGC QSCRW YDLY+ E +G++P Y+MS D+ +I IP LIE

Sbjct: 197 MTARDSNRGGCCQSCRWDYDLYEQQDSAEIPLFAEGDVP--YTMSPKDLNLIQAIPQLIE 254

Query: 245 NGVDSLKIEGRMKSIHYVSTVTNICYKAAVGAYMESPEAFYAIKEELIDELWKVAQRELAT 304

G+DSLK+EGRMKSIHYV+TVT+ Y+ + AY P+ F IK E ++EL K A R+ A

Sbjct: 255 AGIDSLKVEGRMKSIHYVATVTSVYRKVIDAYCSDPDFN-KIKREWLEELEKCANRDFAP 313

Query: 305 GFYYGIPTENEQLFGARRKIPQYKFVGEVVAFDSASMTATIRQNRNIMEGDRIECYGPGEF 364

F+ G PT EQ++G K +Y FVG V+ ++ + T++QRN +GD +E +GP

Sbjct: 314 QFFEGTPTYKEQMYGIHPKRTKYDFVGLVDYNEKTGIVTLQQRNHFQKGDEVEFFGPEI 373

Query: 365 RHFETVVKDLHDADGQKIDRAPNPMELLTISLPREVKPGDMIR 407

F V+ + D DG ++D A +P++++ + ++V P +M+R

Sbjct: 374 NRFTQTVEKIWDEEDGNELDAARHPLQIVKFKVDQKVYPQNMMR 416

An alignment of the GAS and GBS proteins is shown below:

Identities = 386/427 (90%), Positives = 404/427 (94%)

Query: 1 MSNVKKRPEVLSPAGTLEKLKVAIDYGADAVFVGQAYGLRSRAGNFSMEELQEGINIAH 60

MS++KKRPEVLSPAGTLEKLKVAIDYGADAVFVGQAYGLRSRAGNFSMEELQEGI+YAH

Sbjct: 1 MSHMKRPEVLSPAGTLEKLKVAIDYGADAVFVGQAYGLRSRAGNFSMEELQEGIDYAH 60

Query: 61 ARDAKVVAANMVTHEGNELGAGPWFRELDMGLDAVIVSDPALIVICATEAPGLEIHL 120

AR AKVVAANMVTHEGNE+GAG WFR+LRDMGLDAVIVSDPALIVIC+TEAPGLEIHL

Sbjct: 61 ARGAKVVAANMVTHEGNEIGAGEWFRQLRDMGLDAVIVSDPALIVICSTEAPGLEIHL 120

Query: 121 TQASSTNYETFEFWKEMGLTRVVLAREVTMAELAEIRKRTDVEIEAFVHGAMCISYSGRC 180

TQASSTNYETFEFWK MGLTRVVLAREV MAELAEIRKRTDVEIEAFVHGAMCISYSGRC

Sbjct: 121 TQASSTNYETFEFWKAMGLTRVVLAREVNMAELAEIRKRTDVEIEAFVHGAMCISYSGRC 180

Query: 181 VLSNHMSHRDANRGGCSQSCRWKYDLYDMPFGQERQSLKGEIPEPFMSAVDMCMIEHIP 240

VLSNHMSHRDANRGGCSQSCRWKYDLYDMPFG ER+SLKGEIPE +SMS+VDMCMI+HIP

Sbjct: 181 VLSNHMSHRDANRGGCSQSCRWKYDLYDMPFGERRSLKGEIPEDYSMSVDMCMIDHIP 240

Query: 241 DMIENGVDLSKIEGRMKSIHYVSTVTNICYKAAVDAYMESPEAFEAIEDLIDELWKVAQR 300

D+IENGVDLSKIEGRMKSIHYVSTVTNICYKAAV AYMESPEAF AIKE+LIDELWKVAQR

Sbjct: 241 DLIENGVDLSKIEGRMKSIHYVSTVTNICYKAAVGAYMESPEAFYAIKEELIDELWKVAQR 300

Query: 301 ELATGFYYHTPTENEQLFGARRKIPQYKFVGEVVSFDNAKMEATIRQNRNIMEGDRVEFY 360

ELATGFYY PTENEQLFGARRKIPQYKFVGEVV+FD+A M ATIRQNRNIMEGDR+E Y

Sbjct: 301 ELATGFYYGIPTENEQLFGARRKIPQYKFVGEVVAFDSASMTATIRQNRNIMEGDRIECY 360

Query: 361 GPGFRHFECFIDGLRDAEGNKIDRAPNPMELLTITLPNPVKKGDMIRACKEGLVNLYQND 420

GPGFRHFE + L DA+G KIDRAPNPMELLTI+LP VK GDMIRACKEGLVNLYQ D

Sbjct: 361 GPGFRHFETVVKDLHDADGQKIDRAPNPMELLTISLPREVKPGDMIRACKEGLVNLYQKD 420

Query: 421 GTSKTVR 427

-429-

GTSKTVR

Sbjct: 421 GTSKTVR 427

SEQ ID 1094 (GBS385) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 3; MW 50kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 7; MW 75.7kDa).

The GBS385-GST fusion product was purified (Figure 213, lane 7) and used to immunise mice. The resulting antiserum was used for FACS (Figure 312), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 339

A DNA sequence (GBSx0369) was identified in *S.agalactiae* <SEQ ID 1097> which encodes the amino acid sequence <SEQ ID 1098>. This protein is predicted to be collagenase. Analysis of this protein sequence reveals the following:

Possible site: 43

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2208(Affirmative) &lt; succ&gt;

bacterial membrane --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database:

&gt;GP:CAB14677 GB:Z99117 similar to protease [Bacillus subtilis]

Identities = 92/304 (30%), Positives = 161/304 (52%), Gaps = 5/304 (1%)

Query: 1 MEKIILTATAESIEQVKQLLAIGIDRIYVGEENYGLRLPHSFSDDELREIAKLVDHAGKE 60  
M+K L T S + L+ G VGE+ YGLRL FS +++ + ++ H G +

Sbjct: 1 MKKPELIVTPTSTADILPLIQAGATAFLVGEQRYGLRLAGEFSREDVTKAVEIAHKEGAK 60

Query: 61 LTVACNALMHQEMMDNIKPFLELMKEINVDYLVVGDAVGFYINKRDGYNFKLIYDTSVFV 120  
+ VA NA+ H + + + +L + E VD V GD V + + KL + T

Sbjct: 61 VVAVAVNAIFHNDKVGELGEYLAFLEAGVDAAVFGDPAVLMAARESAPDLKLHWSTETT 120

Query: 121 TSSRQVNFQWQHGA VETVLAREIPSEELFKMSENLEFFAEILVYGASVIHHSKRPLLQNY 180  
T+ N+WG+ GA +VLARE+ + + ++ EN E EI V+G + + SKR L+ NY

Sbjct: 121 TNYTTCNYWGRGAARSVLARELNMD SIVEIKENAEVEIEIQVHGTCMFQSKRSLIGNY 180

Query: 181 YNF---THITDEKTRERGLFLAEPGD PESHYSIYEDKHGTHIFINNDINMMTKVTELVEH 237  
+ + + K +E G+FL + + ++ Y I+ED++GTHI ND+ ++ ++ EL++

Sbjct: 181 FEYQKGKVM DIERKKKESGMFLHDK-ERDNKYPIFEDENGTHIMSPNDVCIIDELERLIDA 239

Query: 238 HFTHWKLDGIYCPGDNFVAIAEIFVETARL- IENGTFQDQAF LFDERIRKLHPKGRGLD 296  
+K+DG+ + + + +++ E L +EN + + + ERI + P R +D

Sbjct: 240 GIDSFKIDGVLKMP EYLIEVT KMYREIDL CVENRDEYEAKKEDWIERIESIQPVNRKID 299

Query: 297 TGFY 300

TGF+

Sbjct: 300 TGFF 303

A related GBS nucleic acid sequence <SEQ ID 10949> which encodes amino acid sequence <SEQ ID 10950> was also identified.

-430-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1099> which encodes the amino acid sequence <SEQ ID 1100>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1716(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 245/308 (79%), Positives = 273/308 (88%)

```

Query: 1  MEKIILTATAESIEQVKQLLAIGIDRIYVGEENYGLRLPHSFSDDELREIAKLVDHAGKE 60
          MEKII+TATAESIEQVK LLA G+DRIYVGE NYGLRLPH+FS DELR+IAKLVDHAGKE
Sbjct: 1  MEKIIITATAESIEQVKALLAAGVDRIYVGEANYGLRLPHNFSYDEL RQIAKLVDHAGKE 60

Query: 61  LTVCACNALMHQEMMDNIKPFLELMKEINVDYLVVGDAGVIFYINKRDGYNFKLIYDTSVFV 120
          LTVCACNALMHQ+MMD IKPFL+LM EI VDYL VVG DAGVIFY+NKRDGYNFKLIYDTSVFV
Sbjct: 61  LTVCACNALMHQDMDQIKPFLDLMIETAVDYLVVGDAGVIFYVKNKRDGYNFKLIYDTSVFV 120

Query: 121 TSSRQVNFWGQHGAVETVLAREIPSEELFKMSENLEFP AEILVYGASVIHHSKRPLLQNY 180
          TSSRQVNFWGQHGAVE+VLAREIPS ELF ++ENLEFP AE+LVYGASVIHHSKRPLL+NY
Sbjct: 121 TSSRQVNFWGQHGAVESVLAREIPS AELFTLAENLEFP AEVLVYGASVIHHSKRPLENY 180

Query: 181 YNFTHITDEKTRERGLFLAEPGD PESHYSIYEDKHGTHIFINNDINMMTKVTELVEHHFT 240
          Y+FT I DE +RERGLFLAEPGD SHYSIYED HGTHIFINNDI+MM+K+ EL H T
Sbjct: 181 YHFTKIDDEVSRERGLFLAEPGD ASSHSIYEDNHGTHIFINNDIDMMSKLGELYAHGLT 240

Query: 241 HWKLDGIYCPGDNFVAIAEIVFETARLIENGTF TQDQAFLFDERIRKLHPKGRGLDTGFY 300
          HWKLDGIYCPGD+FVAI ++F++ L+E G FTQ++A D+ + HP GRGLDTGFY
Sbjct: 241 HWKLDGIYCPGD D FVAITKLF IQAKT LLEAGQFTQEAEKLDQAVHAHPAGRGLDTGFY 300

Query: 301 DFD PSTVK 308
          +FDP TVK
Sbjct: 301 EFD PKTVK 308

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 340

A DNA sequence (GBSx0371) was identified in *S.agalactiae* <SEQ ID 1101> which encodes the amino acid sequence <SEQ ID 1102>. This protein is predicted to be cDNA EST yk542c12.5 comes from this gene. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP: AAD15622 GB:U75480 unknown [Streptococcus mutans]

Identities = 69/152 (45%), Positives = 101/152 (66%), Gaps = 12/152 (7%)

```

Query: 1  MSKLFKTLVISAASGAAAYFLTTKKGKELRNKAEKFYGEYKENPEEYHQIAKDKASEYS 60
          MSK KT +I A +GAAAYFL+T KGK+ +K + + +YKENP+EYHQ A DK +EY
Sbjct: 1  MSKFLKTAIIGAGTGAAAYFLSTDGKGQFKKKIHQTFTDYKENPKEYHQYAADKVNEYK 60

```



-431-

Query: 61 NLAVDTFKDYKGFESGELTTEDIVSAVKEKSGEVVDVFANDFVNQAKSKFSDEDTAKKED 120  
 ++AV +FKDYK KFE+GELT ++I+S+VKEK+ + FAN ++Q K + T +K +  
 Sbjct: 61 DVAVHSFKDYKDKFETGELTKDNIISVKEKASQAGKFANSKLSQVKDHLA--QTVEKAE 118

Query: 121 KAP-----ETKVEDIVIDYKENTEDKE 142  
 + +V+DIVIDY+ + K+  
 Sbjct: 119 ASTNDAGIPLGEMKAQVDDIVIDYQAEKTKK 150

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1103> which encodes the amino acid sequence <SEQ ID 1104>. Analysis of this protein sequence reveals the following:

Possible site: 26  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.81 Transmembrane 15 - 31 ( 14 - 31)

15 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1723(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 20 A related sequence was also identified in GAS <SEQ ID 9117> which encodes the amino acid sequence <SEQ ID 9118>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 19  
 >>> Seems to have a cleavable N-term signal seq.

25 ----- Final Results -----  
 bacterial outside --- Certainty= 0.300(Affirmative) < succ>  
 bacterial membrane --- Certainty= 0.000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

- 30 An alignment of the GAS and GBS proteins is shown below:

Identities = 69/140 (49%), Positives = 91/140 (64%), Gaps = 8/140 (5%)

Query: 1 MSKLFKTLVISAASGAAAYFLTTKKGKELRKNAEKFYGEYKENPEEYHQIAKDKASEYS 60  
 M+K FK LVI A SG AAAYFL+T+KGK L+ AEK Y YKE+P++YHQ AK+K SEYS  
 Sbjct: 8 MNKSFKNLVIGAVSGVAAAYFLSTKKGKALKNRAEKAYQAYKESPDYHQAFAKEKGSEYS 67

Query: 61 NLAVDTFKDYKGFESGELTTEDIVSAVKEKSGEVVDVFANDFVNQAKSKFSDEDTAKKE 119  
 +LA DTF D K K SG+LT ED++ +K+K+ FV + K ++ E K++  
 Sbjct: 68 HLRDRTFYDVKDKLASGDLTKEDMLDLLKDKT-----TAFVQTKKETLAEVEAKEKQD 120

Query: 120 DKAPETKVEDIVIDYKENTE 139  
 D + EDI+IDY E E  
 Sbjct: 121 DVIIDLNEEDIIIDYTEQDE 140

- 45 SEQ ID 1102 (GBS164) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 30 (lane 4; MW 17.4kDa).

The GBS164-His fusion product was purified (Figure 115A; see also Figure 200, lane 4) and used to immunise mice (lane 1+2+3 product; 20µg/mouse). The resulting antiserum was used for Western blot, FACS (Figure 115B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-432-

**Example 341**

A DNA sequence (GBSx0372) was identified in *S.agalactiae* <SEQ ID 1105> which encodes the amino acid sequence <SEQ ID 1106>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
5  >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood =-16.93    Transmembrane    6 - 22 ( 1 - 31)

----- Final Results -----
10      bacterial membrane --- Certainty=0.7771(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15  >GP:AAD15621 GB:U75480 unknown [Streptococcus mutans]
    Identities = 88/129 (68%), Positives = 112/129 (86%)

Query: 1  MIEIAVLIIAIAFVVLVLGILFVLKKVSETIEETKQTIKVLTSDEVNVTLYQTNEILAKAN 60
          M EIA+LI+AI AF VLV+ ++ +L+K+S+T++E++QT+K+LTSDVNVTLYQTNE+LAKAN
Sbjct: 1  MWEIALLIVAIAFAVLVIYLILLRKRISDTVDESRTLKILTSDVNVTLYQTNEILLAKAN 60

20  Query: 61  VLVDVNGKVSTIDPLFVAIADLSSESVSDLNLQARHIGQKASSATSSVTKAGSALAIGKA 120
          VLV+DVNGKV TIDPLF AIADLS SVSDLN QAR+ G+K +T++V KAG+A GK
Sbjct: 61  VLVEDVNGKVETIDPLFTAIAADLSVSVSDLNRQARYFGKKTRKSTANVGKAGAAAYTFGKV 120

25  Query: 121 ASKIFRKKG 129
          ASK+FRKKG
Sbjct: 121 ASKLFRKKG 129

```

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1107> which encodes the amino acid sequence <SEQ ID 1108>. Analysis of this protein sequence reveals the following:

```

Possible site: 16
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -0.85    Transmembrane    18 - 34 ( 17 - 34)

35  ----- Final Results -----
      bacterial membrane --- Certainty=0.1341(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

40 The protein has homology with the following sequences in the databases:

```

>GP:AAD15621 GB:U75480 unknown [Streptococcus mutans]
    Identities = 83/128 (64%), Positives = 110/128 (85%)

45  Query: 6  ISLMIIALAFVALVIFLIIVLKKVSETIDEAKKTIISVLTSDEVNVTLHQTNDILAKANILV 65
          I+L+I+A+AF LVI+LI++L+K+S+T+DE+++T+ +LTSDVNVTL+QTN++LAKAN+LV
Sbjct: 4  IALLIVAIAFAVLVIYLILLRKRISDTVDESRTLKILTSDVNVTLYQTNEILLAKANVLV 63

Query: 66  EDVNGKVATIDPLFVAIADLSSESLSDLSQARHFGQKATNATGNVSKAGKLALVGKVASK 125
          EDVNGKV TIDPLF AIADLS S+SDLN QAR+FG+K +T NV KAG GKVASK
50  Sbjct: 64  EDVNGKVETIDPLFTAIAADLSVSVSDLNRQARYFGKKTRKSTANVGKAGAAAYTFGKVASK 123

Query: 126 VFGKKGEK 133
          +F KKG++
Sbjct: 124 LFRKKGKQ 131

```

55 An alignment of the GAS and GBS proteins is shown below:

```

    Identities = 92/131 (70%), Positives = 116/131 (88%)

60  Query: 1  MIEIAVLIIAIAFVVLVLGILFVLKKVSETIEETKQTIKVLTSDEVNVTLYQTNEILAKAN 60
          ++ I+++IIA+AFV LV+ ++ VLKKVSETI+E K+TI VLTSDEVNVTL+QTN+ILAKAN

```

-433-

Sbjct: 3 LVGISLMIIALAFVALVIFLIIVLKKVSETIDEAKKTISVLTSDVNVTLHQTNLILAKAN 62

Query: 61 VLVDVNGKVSTIDPLFVAIADLSESVSDLNLQARHIGQKASSATSSVTKAGSALAIGKA 120  
+LV+DVNGKV+TIDPLFVAIADLSE+SDLN QARH GQKA++AT +V+KAG +GK

5 Sbjct: 63 ILVEDVNGKVATIDPLFVAIADLSESLSDLNSQARHFGQKATNATGNVSKAGKLALVGKV 122

Query: 121 ASKIFRKKGDK 131  
ASK+F KKG+K

10 Sbjct: 123 ASKVFGKKGEK 133

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 342

A DNA sequence (GBSx0373) was identified in *S.agalactiae* <SEQ ID 1109> which encodes the amino acid sequence <SEQ ID 1110>. Analysis of this protein sequence reveals the following:

Possible site: 13  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.0462(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 343

A DNA sequence (GBSx0374) was identified in *S.agalactiae* <SEQ ID 1111> which encodes the amino acid sequence <SEQ ID 1112>. This protein is predicted to be prolipoprotein diacylglycerol transferase (lgt). Analysis of this protein sequence reveals the following:

Possible site: 29  
>>> Seems to have an uncleavable N-term signal seq

35 INTEGRAL Likelihood = -8.39 Transmembrane 231 - 247 ( 225 - 251)  
INTEGRAL Likelihood = -7.64 Transmembrane 89 - 105 ( 87 - 107)  
INTEGRAL Likelihood = -5.20 Transmembrane 18 - 34 ( 13 - 36)  
INTEGRAL Likelihood = -1.86 Transmembrane 46 - 62 ( 46 - 64)

----- Final Results -----

40 bacterial membrane --- Certainty=0.4354(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9721> which encodes amino acid sequence <SEQ ID 9722> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC80171 GB:U75480 putative prolipoprotein diacylglycerol  
transferase [*Streptococcus mutans*] (ver 3)  
Identities = 184/257 (71%), Positives = 226/257 (87%)

50 Query: 2 MINPVAIRLGPFSIRWYAICIVSGMLLAVYLAMKEAPRKNIKSDDILDFILMAFPLSIVG 61  
MINP+AI+LGP +IRWY+ICIV+G++LAVYL ++EAP+KNIKSDD+LDFIL+AFPL+IVG

-434-

Sbjct: 1 MINPIAIKLGPLTIRWYSICIVTGLILAVYLTIREAPKKNIKSDDVLDFILIAFPLAIVG 60

Query: 62 ARIYYVIFEWAYYSKHPVEIIAIWNGGIAIYGGLITGAILLVIFSYRRLINPIDFLDIAA 121  
AR+YYVIF+W YY K+P EI IW+GGIAIYGGL+TGA++L IFSY R+I PIDFLD+AA

5 Sbjct: 61 ARLYYVIFDWDYLLKNPSEIPVIWHGGIAIYGGLLTGALVLFIFSYRMIKPIDFLDVAA 120

Query: 122 PGVMIAQAIGRWGNFINQEAYGRAVKNLNYPNFIKNQMYIDGAYRVPTFLYESLWNFLG 181  
PGVM+AQ+IGRWGNF+NQEAYG+ V LNY+P+FI+ QMYIDG YR PTFLYESLWN LG

10 Sbjct: 121 PGVMLAQSIGRWGNFVNQEAYGKTVTQLNLYLPDFIRKQMYIDGHYRTPPTFLYESLWNLLG 180

Query: 182 FVIIMSIRHRPRTLKQGEVACFYLVWYGCGRFIIIEGMRTDSLILAGLRVSQWLSVILVII 241  
F+IIM +R RP LK+GEVA FYL+WYG GRF+IEGMRTDSL A LRVSQWLSV+LV++

Sbjct: 181 FIIIMILRRRNPNLLKEGEVAFFYLIWYGSGRFVIEGMRTDSLMFASLRVSQWLSVLLVVV 240

15 Query: 242 GIVMIIYRRREQHISYY 258  
G++++ RRR I YY

Sbjct: 241 GVILMVIRRRNHAIPYY 257

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1113> which encodes the amino acid  
20 sequence <SEQ ID 1114>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have an uncleavable N-term signal seq

25 INTEGRAL Likelihood = -7.01 Transmembrane 229 - 245 ( 222 - 249)  
INTEGRAL Likelihood = -6.90 Transmembrane 45 - 61 ( 40 - 68)  
INTEGRAL Likelihood = -4.41 Transmembrane 17 - 33 ( 11 - 35)  
INTEGRAL Likelihood = -4.14 Transmembrane 87 - 103 ( 86 - 106)  
INTEGRAL Likelihood = -0.27 Transmembrane 170 - 186 ( 170 - 186)

30 ----- Final Results -----  
bacterial membrane --- Certainty=0.3803(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

35 >GP:AAC80171 GB:U75480 putative prolipoprotein diacylglycerol  
transferase [Streptococcus mutans] (ver 3)  
Identities = 176/258 (68%), Positives = 217/258 (83%)

40 Query: 1 MINPIALKCGPLAIHWHYALCILSGLVLAVYLASKEAPKKGISSDAIFDFILIAFPLAIVG 60  
MINPIA+K GPL I WY++CI++GL+LAVYL +EAPK I SD + DFILIAFPLAIVG  
Sbjct: 1 MINPIAIKLGPLTIRWYSICIVTGLILAVYLTIREAPKKNIKSDDVLDFILIAFPLAIVG 60

45 Query: 61 ARIYYVIFEWSSYYVKHLDEIIAIWNGGIAIYGGLITGALVLLAYCYNKVLNPIHFLDIAA 120  
AR+YYVIF+W YY+K+ EI IW+GGIAIYGGL+TGA+L + Y +++ PI FLD+AA  
Sbjct: 61 ARLYYVIFDWDYLLKNPSEIPVIWHGGIAIYGGLLTGALVLFIFSYRMIKPIDFLDVAA 120

50 Query: 121 PSMVMAQAIGRWGNFINQEAYGKAVSQLNLYLPSFIQKQMFIEGYSYRIPTFLYESLWNLLG 180  
P VM+AQ+IGRWGNF+NQEAYGK V+QLNLYL FI+KQ+I+G YR PTFLYESLWNLLG  
Sbjct: 121 PGVMLAQSIGRWGNFVNQEAYGKTVTQLNLYLPDFIRKQMYIDGHYRTPPTFLYESLWNLLG 180

55 Query: 181 FVIIMMWRKPKSLLDGEIFAFLYLIWYGSGRLVIEGMRTDSLMFLGIRISQYVSALLIII 240  
F+IIM+ RR+P L +GE+ FYLIWYGSGR VIEGMRTDSLMF +R+SQ++S LL+++  
Sbjct: 181 FIIIMILRRRNPNLLKEGEVAFFYLIWYGSGRFVIEGMRTDSLMFASLRVSQWLSVLLVVV 240

Query: 241 GLIFVIKRRRQKGISYYQ 258  
G+I ++ RRR I YYQ  
Sbjct: 241 GVILMVIRRRNHAIPYYQ 258

An alignment of the GAS and GBS proteins is shown below:

60 Identities = 176/257 (68%), Positives = 221/257 (85%)

Query: 2 MINPVAIRLGPFISIRWYAICIVSGMLLAVYLAKEAPRKNIKSDDIILDFILMAFPLSIVG 61  
MINP+A++ GP +I WYA+CI+SG++LAVYLA KEAP+K I SD I DFIL+AFPL+IVG  
Sbjct: 1 MINPIALKCGPLAIHWHYALCILSGLVLAVYLASKEAPKKGISSDAIFDFILIAFPLAIVG 60

5

```
Query: 62  ARIYYVIFEWAYYSKHPVEIIAIWNGGIAIYGGITGAILLVIFSYRRLINPIDFLDIAA 121
           ARIYYVIFEW+YY KH  EIIAIWNGGIAIYGGITGA++L+ + Y +++NPI FLDIAA
Sbjct: 61  ARIYYVIFEWSYVVKHLDEIIAIWNGGIAIYGGITGALVLLAYCYNKVLNPIHFLDIAA 120

Query: 122  PGMVIAQAIGRWGNFNINQAYGRAVKNLNYPNFIKNQMYIDGAYRVPTFLYESLWNFLG 181
           P VM+AQAIGRWGNFNINQAYG+AV  LNY+P+FI+ QM+I+G+YR+PTFLYESLWN LG
Sbjct: 121  PSMVMAQAIGRWGNFNINQAYGKAVSQLNLYLPSFIQKQMFIEGSYRIPTFLYESLWNLLG 180

10
Query: 182  FVIIMSIRHRPRTLKQGEVACFYLVWYGCGRFIIEGMRTDSL YLAGLRVSQWLSVILVII 241
           FVIIM R +P++L GE+ FYL+WYG GR +IEGMRTDSL G+R+SQ++S +L+II
Sbjct: 181  FVIIMMWRKPKSLLDGEIFAFYLIWYSGRLVIEGMRTDSL MFLGIRISQYVSALLIII 240

Query: 242  GIVMIIYRRREQHISYY 258
           G++ +I RRR++ ISYY
Sbjct: 241  GLIFVIKRRRQKGISYY 257

15
```

A related GBS gene <SEQ ID 8557> and protein <SEQ ID 8558> were also identified. Analysis of this protein sequence reveals the following:

```

20      Lipop: Possible site: -1      Crend: 0
      McG: Discrim Score:      2.45
      GvH: Signal Score (-7.5): -2.9
            Possible site: 39
      >>> Seems to have an uncleavable N-term signal seq
25      ALOM program      count: 3 value: -8.39 threshold: 0.0
            INTEGRAL      Likelihood = -8.39      Transmembrane      209 - 225 ( 203 - 229)
            INTEGRAL      Likelihood = -7.64      Transmembrane      67 - 83 ( 65 - 85)
            INTEGRAL      Likelihood = -1.86      Transmembrane      24 - 40 ( 24 - 42)
            PERIPHERAL      Likelihood = 0.79      92
30      modified ALOM score:      2.18

      *** Reasoning Step: 3

      ----- Final Results -----
35      bacterial membrane --- Certainty=0.4354(Affirmative) < succ>
            bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

[illegible]

5  
 918 948 978 1008 1038 1068 1098 1128  
 RTDSL<sup>Y</sup>LAGLRVSQWLSVILVIIGIVMIIYRRREQHISY\*<sup>T</sup>EEVL\*<sup>K</sup>LLY\*<sup>L</sup>LLPLRLLF\*<sup>F</sup>EYFSF\*<sup>K</sup>KYQKRLRKP  
 ||||| :| ||||| ||| :|| ::::: ||| :| ||  
 RTDSL<sup>M</sup>FASLRVSQWLSVLLVVVGIVLMVIRRRNHAI<sup>P</sup>YYQC  
 230 240 250

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 344

10 A DNA sequence (GBSx0375) was identified in *S.agalactiae* <SEQ ID 1115> which encodes the amino acid sequence <SEQ ID 1116>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence
```

```

15      ----- Final Results -----
           bacterial cytoplasm --- Certainty=0.2817(Affirmative) < succ>
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

20 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA77782 GB:AB027460 Hpr kinase [Streptococcus bovis]  
Identities = 264/309 (85%), Positives = 292/309 (94%)

25 Query: 1 MAVTVQMLVDR LKLNVIYGDEHLLSKRITTADISRPGLEMTGYFDYYAPERLQLVGMKEW 60  
M+VTV+MLVD++KL+VIYGD+ LLSK ITT+DISRPGLEMTGYFDYY+PERLQL+GMKEW  
Sbjct: 1 MSVTVKMLVDKVKLDVIYGD+LLSKEITTS+DISRPGLEMTGYFDYYSPERLQL+GMKEW 60

```

Query: 61 SYLMAMTGHNRYQVLREMFQKETPAIVVARDLEIPEEMYEAAKDTGIALIQSKAPTSRLS 120
          SYL MT HNR VLREM + ETPAI+VAR+L IPEEM AAK+ GIALIQS PTSRLS
30 Subject: 61 SYLTKMTSHNRRHVLRMIKPETPAIIVARNLAIPEEMISAAKEKGIALIQSHVPTSRLS 120

```

Query: 121 GEVSWYLDSCLAERTSVHGVLMIDIYGMGVLIQDGSIGKSETGLELVKRGHRLVADDRVD 180  
 GE+SWYLDSCLAERTSVHGVLMIDIYGMGVLIQDGSIGKSETGLELVKRGHRLVADDRVD  
 Subject: 121 GEMSWYLDSCLAERTSVHGVLMIDIYGMGVLIQDGSIGKSETGLELVKRGHRLVADDRVD 180

Query: 181 VYAKDEETLWGEPAEILRHLLLEIRGVGIIDIMSLYGASAVKDSSQVQLAITYLENFETGKV 240  
V+AKDEETLWGEPAEILRHLLLEIRGVGIID+MSLYGASAVKDSSQVQLAITYLEN+E+GKV  
Sbjct: 181 VFAKDEETLWGEPAEILRHLLLEIRGVGIIDVMSLYGASAVKDSSQVQLAITYLENYESGKV 240

40      Query: 241 FDRLGNGNEEIELSGVKVPRIRIPVKITGRNVS SVVIEAAAMNHRAKQMGF DATQTTFEDRLT 300  
                  FDRLGNGNEE+ELSGVK+PR+RIPV+TGRN+SVVIEAAAMN+RAKQMGF DAT+TFE+RLT  
      Subject: 241 FDRLGNGNEEIELSGVKIPRLRIPVOTGRNMSV SVVIEAAAMNRAKQMGF DATKTTFEERLT 300

45           Query: 301 HLISQNEVN 309  
                      LI++NE N  
          Sbjct: 301 QLITKNEGN 309

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1117> which encodes the amino acid sequence <SEQ ID 1118>. Analysis of this protein sequence reveals the following:

```
50      Possible site: 13
      >>> Seems to have no N-terminal signal sequence
```

----- Final Results -----

55           bacterial cytoplasm --- Certainty=0.2391(Affirmative) < succ>  
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 255/309 (82%), Positives = 288/309 (92%)

-437-

Query: 1 MAVTVQMLVDRLKLNVIYGDHLLSKRITTADISRPGLEMTGYFDYYAPERLQLVGMKEW 60  
 M VTV+MLV ++KL+V+Y ++LLSK ITT+DISRPGLEMTGYFDYYAPERLQL GMKEW  
 Sbjet: 32 MTVTVKMLVQVKLDVVYATDNLLSKEITTSDISRPGLEMTGYFDYYAPERLQLFGMKEW 91

Query: 61 SYLMAMTGHNRYQVLREMFQKETPAIVVARDLEIPEEMYAAKDTGIAILQSKAPTSRLS 120  
 SYL MT HNRY VL+EMF+K+TPA+VV+R+L IP+EM +AAK+ GI++L S+ TSRL+  
 Sbjet: 92 SYLTQMTSHNRYSVLKEMFKKDTPAVVVSRNLAIPKEMVQAAKEEGISLLSSRVSTSRLA 151

Query: 121 GEVSWYLDSCLAERTSVHGVLMIDIYGMGVLIQGD SGIGKSETGLELVKRGHRLVADDRVD 180  
 GE+S++LD+ LAERTSVHGVLMIDIYGMGVLIQGD SGIGKSETGLELVKRGHRLVADDRVD  
 Sbjet: 152 GEMSYFLDASLAERTSVHGVLMIDIYGMGVLIQGD SGIGKSETGLELVKRGHRLVADDRVD 211

Query: 181 VYAKDEETLWGEPAEILRHLLLEIRGVGIIDIMSLYGASAVKDSSQVQLAIYLENFETGKV 240  
 VYAKDEETLWGEPAEILRHLLLEIRGVGIID+MSLYGASAVKDSSQVQLAIYLENFE GKV  
 Sbjet: 212 VYAKDEETLWGEPAEILRHLLLEIRGVGIIDVMSLYGASAVKDSSQVQLAIYLENFEAGKV 271

Query: 241 FDR LGNGNEEIELSGVKVPRIRIPVKTGRNVSVVIEAAAMNHRAKMGFDATQTFEDRLT 300  
 FDR LGNGNEEI SGV++PRIRIPVKTGRNVSVVIEAAAMNHRAK+MGFDAT+TFEDRLT  
 Sbjet: 272 FDR LGNGNEEITFSGVRIPRIRIPVKTGRNVSVVIEAAAMNHRAKMGFDATKTFEDRLT 331

Query: 301 HLISQNEVN 309  
 LI++NEV+  
 Sbjet: 332 QLITKNEVS 340

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 345

A DNA sequence (GBSx0376) was identified in *S.agalactiae* <SEQ ID 1119> which encodes the amino acid sequence <SEQ ID 1120>. Analysis of this protein sequence reveals the following:

Possible site: 28  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1836(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9719> which encodes amino acid sequence <SEQ ID 9720> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 346

A DNA sequence (GBSx0377) was identified in *S.agalactiae* <SEQ ID 1121> which encodes the amino acid sequence <SEQ ID 1122>. Analysis of this protein sequence reveals the following:

Possible site: 37  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -4.88 Transmembrane 35 - 51 ( 31 - 59)

----- Final Results -----

bacterial membrane --- Certainty=0.2954(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-438-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC67275 GB:AF017113 Yv1C [Bacillus subtilis]

Identities = 21/63 (33%), Positives = 36/63 (56%), Gaps = 2/63 (3%)

Query: 3 SSFYKQQRKGLVCGVVAGLADKYNWDLALSRVLIALILYFTKF--GLLLYILLAVFLPYK 60

+ Y+ K K + GV+ GLA+ +NWD +L RV+ ++ T LL+YI+ +P +

Sbjct: 2 NKLYRSEKNKKIAGVIGGLAEYFNWDASLLRVITVILAIMTSVLPVLLIYIWIWIFIVPSE 61

Query: 61 EDI 63

D+

Sbjct: 62 RDM 64

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1123> which encodes the amino acid sequence <SEQ ID 1124>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -5.26 Transmembrane 39 - 55 ( 31 - 61)

----- Final Results -----

bacterial membrane --- Certainty=0.3102(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 60/90 (66%), Positives = 77/90 (84%), Gaps = 3/90 (3%)

Query: 1 MKSSFYKQQRKGLVCGVVAGLADKYNWDLALSRVLIALILYFTKFGLLLYILLAVFLPYK 60

+++ FYKQRK +LV GV+AGLADKY WDLAL+RVL AL++Y T FG+LLYILLA+FLPYK

Sbjct: 1 VETKFYKQQRKRLVAGVIAGLADKYGWDLALARVLAALLIYGTGFGVLLYILLAIFLPYK 60

Query: 61 EDIIETR-RQGPRRRKDAEPV--DDDGWFW 87

ED++E R +GPRRRKDA+ + ++DGWFW

Sbjct: 61 EDLLEERYGRGPRRRKDADVLNEEDGWFW 90

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 347

A DNA sequence (GBSx0378) was identified in *S.agalactiae* <SEQ ID 1125> which encodes the amino acid sequence <SEQ ID 1126>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3577(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9717> which encodes amino acid sequence <SEQ ID 9718> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04250 GB:AP001508 unknown conserved protein [Bacillus halodurans]

Identities = 379/729 (51%), Positives = 515/729 (69%), Gaps = 25/729 (3%)

Query: 29 ENLNITQIAIDLGIKASQIEKVLLELTDEGNTIPFIARYRKEMTGNLDEVQIKSIIDLKDS 88



-439-

E I +A +L +K + I++V++L EGNT+PFIARYRKE+TG +DEV+I+ + +  
 Sbjct: 8 EEHTIKTLAKELSLKPNYIKQVIQLLHEGNTVPFIARYRKELTGGMDEVKIREVSEKWTY 67  
 Query: 89 MTALSDRKTTLAKIEEQGKLTQELKKAIEEATKLADVEELYLPYKEKRRTKATTAREAG 148  
 L +RK V+ +EEQGKLT E KK +E+A KL +VE+LY PYK+KRRT+AT+A+E G  
 Sbjct: 68 ANQLHERKEEVIRLVEEQGKLTDEWKKTVEQAQKLQEVEDLYRPPYKQKRRTATVAKEKG 127  
 Query: 149 LFPLARLI--LQNKDNLEEEAQNLYLTGDFETTT--KALSGAVDILIEAFSEDNKLRSWTY 204  
 L PLA + L + +EA+ YL+ E T L GA DI+ E ++D LR  
 Sbjct: 128 LEPLAEWLFSLPRDGLQEAQVYLSVEHELTKVEDVLQGAQDIIAEWIADDADLRKRIR 187  
 Query: 205 NEIWNYSITAVVKDESLEKQVFKIYYDFSEKISKLHGYQVLALNRGEKMGVLKVNFEH 264  
 + + S+ A VK E LDEK V+++YYD+ E + L ++ LALNRGEK VL+V  
 Sbjct: 188 SLGFKEGSVIAKVKKEELDEKGVYEMYYDYEEPVRTLVPHRTLALNRGEKEDVLRVTIRF 247  
 Query: 265 NLEKMFRTF----FAVRFKETS-QYIDDLIVQTVKKKIVPAMERRIRTELSEGAEDGAISL 319  
 ++++ F RF + Y+ I K+ I P++ER IR EL+E AE+ AI +  
 Sbjct: 248 PVDRIIEMSEKTFIRRFSGPAVPYVKAIEDGYKRLIEPSIEREIRHELTEKAEQAIHI 307  
 Query: 320 FSENLRNLLVSPKGMVLGFDPAFRGTAKLAVVDQTGKLMTTQVIYPVPPANQAKIEQ 379  
 F+ENLR+LLL P+KGK+VLG DPA+RTG KLA+VD+TGK++ QVIYP PP N+ +  
 Sbjct: 308 FAENLRSLLLQPPKIGKVLGLDPAVRTGCKLAIVDETGVLDIQVIYPTPPKNE--VAA 365  
 Query: 380 SKIELAKLIKEFNIEIIAIGNGTASRESEAFVAEVLQDFPD-VSYVIVNESGASVYSASE 438  
 +K + KLI ++ +E+IAIGNGTASRESE F+A++++D P + Y+IVNE+GASVYSASE  
 Sbjct: 366 AKKIVKKLIADYGVEMIAIGNGTASRESEQFIADLIKDLPTIYYLIVNEAGASVYSASE 425  
 Query: 439 LARHEFPDLTVEKRSASISARRLQDPLAELVKIDPKSIGVGQYQHDVSQKLAENLDFVV 498  
 + R EFPDL VE+RSA+SIARRLQDPLAELVKIDPKS+GVGQYQHDVSQK+L E+L FVV  
 Sbjct: 426 IGREEFPDLQVEERSAVSISARRLQDPLAELVKIDPKSVGVGQYQHDVSQKRLNESLTFVV 485  
 Query: 499 ETVVNQGVNVNTASPALLAHVSGLNKTISENIVKYREENGQIKSRAEIKKVPRLGAKAF 558  
 ETVVNQGVNVNTASP+LL +V+GL+KT+++NIVK REE G+ +RA++K +PRLGAK +  
 Sbjct: 486 ETVVNQGVNVNTASPSLLQYVAGLSKTVAKNIVKKREEAGRFTARAQLKDIPRLGAKTY 545  
 Query: 559 EQAAGFLRIPNAKNFLDNTGVHPESYEA VKKLLDQLTIKELD---DLAKEKLQNLDIAT 615  
 EQ GFLRI + N LD T +HPESY+ KLL ++ D + K+KLQ LD+ A  
 Sbjct: 546 EQCIGFLRIMDGNLLDATAIHPEYKVTDKLLSEVGATAADVGIEDLKKKLQALDVSAM 605  
 Query: 616 AESIGVGQETLKDIIEDLLKPRDLRDDFEAPVLRHDVLDVSDLVKGQELQGTVRNVVDF 675  
 A ++ VG TLKD+I+ L++P RD RD+ P+L+ DVL + DL G ELQGTVRNVVDF  
 Sbjct: 606 AATLDVGVP TLKDMIDALIRPTRDPRDEVAKPLLQDVLQLEDLLPGMELQGTVRNVVDF 665  
 Query: 676 GAFVDIGVHEDGLIHQSRLIKRKRDKTRKMPPLQHPKYL SVGDIVTVWVVEVDAERSR 735  
 G FVDIGV +DGL+H S+L R ++HP + ++VG+IVTVWV +VD ++ R  
 Sbjct: 666 GVFVDIGVKQDGLVHISKLANRY-----IKHPLEVVTVGEIVTVWVEDVDIKGR 715  
 Query: 736 IGLSLIKPD 744  
 I L++++P+  
 Sbjct: 716 IALTMLRPE 724

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1127> which encodes the amino acid sequence <SEQ ID 1128>. Analysis of this protein sequence reveals the following:

Possible site: 25  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2207(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 532/716 (74%), Positives = 619/716 (86%), Gaps = 10/716 (1%)  
 Query: 28 MENLNITQIAIDLGKASQIEKVLLELTDEGNTIPFIARYRKEMTGNLDEVQIKSIIDLDK 87

-440-

MEN N IA L + QIE+VL LT +GNTIPFIARYRKE+TGNLDEV IKSIIID+DK  
 Sbjct: 1 MENNNHNHIAEALSVSLHQIEQVLAALTAQGNTIPFIARYRKEVTGNLDEVVIKSIIIDMDK 60

Query: 88 SMTALSDRKTTLAKIEEQGKLTQELKKAIEEATKLADVEELYLPYKEKRRTKATIAREA 147  
 S+T L++RK T+LAKIEEQGKLT +L+ +IE KLAD+EELYLPYKEKRRTKATIAREA  
 Sbjct: 61 SLTTLNERKATILAKIEEQGKLTDLQRTSIEATEKLADLEELYLPYKEKRRTKATIAREA 120

Query: 148 GLFPLARLILQNKDNLEEEAQNLYLTDGFETTTKALSGAVDILIEAFSEDNKLRSWTYNEI 207  
 GLFPLARLILQN NLE A+ ++T+GF + +AL+GAVDIL+EA SED KLRSWTYNEI  
 Sbjct: 121 GLFPLARLILQNAQNLETAABPFVTEGFASPQEALAGAVDILVEAMSEDAKLRSWTYNEI 180

Query: 208 WNYSSITAVVKDESLEKQVFKIYYDFSEKISKLHGYQVLALNRGEKMGVLKVNFEHNLE 267  
 W YS + + +KDE LDEK+VF+IYYDFS+++S + GY+ LALNRGEK+G+LKV+FEHNLE  
 Sbjct: 181 WQYSRLVSTLKDQLEKQVFKIYYDFSQVSNMQGYRTLALNRGEKLGILKVSFEHNLE 240

Query: 268 KMFRRFAVRFKETSQYIDDLIVQTVKKKIVPAMERRIRTELSEGAEDGAISLFSENLRNL 327  
 KM RFF+VRFKET+ YI+++I QT+KKKIVPAMERR+R+ELS+ AEDGAI LFSENLR+L  
 Sbjct: 241 KMQRFFSVRFKETNPYIEEVINQTIKKKIVPAMERRVRSELSDAEDGAIHLSFENLRHL 300

Query: 328 LLVSPLKGMVLGFDPAFRTGAKLAVVDQTGKLMTTQVIYPVPPANQAKIEQSKIELAKL 387  
 LLVSPLKGMVLGFDPAFRTGAKLA+VDQTGKL+TTQVIYPV PA+Q KI+ +K L +L  
 Sbjct: 301 LLVSPLKGMVLGFDPAFRTGAKLAIVDQTGKLTTQVIYPVAPASQTKIQAAKETLTQL 360

Query: 388 IKEFNIEIIAIGNGTASRESEAFVAEVLQDFPDVSYVIVNESGASVYSASELARHEFPDL 447  
 I+ + I+IIAIGNGTASRESEAFVA+VL+DFP+ SYVIVNESGASVYSASELARHEFPDL  
 Sbjct: 361 IETYQIDIIAIGNGTASRESEAFVADVLKDFPNTSYVIVNESGASVYSASELARHEFPDL 420

Query: 448 TVEKRSAISIARRLQDPLAELVKIDPKSIGVGQYQHDVSKKKLAENLDFVETVNVQGV 507  
 TVEKRSAISIARRLQDPLAELVKIDPKSIGVGQYQHDVSKKKL+ENL FVV+TVNVQGV  
 Sbjct: 421 TVEKRSAISIARRLQDPLAELVKIDPKSIGVGQYQHDVSKKKLSENLFVVDTVNVQGV 480

Query: 508 NVNTASPALLAHVSGLNKTI SENIVKYREENGQIKSRAEIKKVPRLGAKAFEQAAGFLRI 567  
 NVNTASP+LLAHVSGLNKTI SENIVKYREENG + SRA+IKKVPRLGAKAFEQAAGFLRI  
 Sbjct: 481 NVNTASPSLLAHVSGLNKTI SENIVKYREENGALTSRADIKKVPRLGAKAFEQAAGFLRI 540

Query: 568 PNAKNFLDNTGVHPESYEAVKKLLDQLTIKELDDLAKELQNLDLIATAESIGVGQETLK 627  
 P AKN LDNTGVHPESY AVK+L L I++LDD AK L + + AE++ +GQETLK  
 Sbjct: 541 PGAKNILDNTGVHPESYPAVKELFKVLGIQDLDDAAKATLAAVQVPQMAETLAIGQETLK 600

Query: 628 DIIEDLLKPGRDLRDDFEAPVLRHVDLVDLKVQELQGTVRNVVDFGAFVDIGVHEDG 687  
 DII DLLKPGRDLRDDFEAP+LR D+LD+ DL++GQ+L+GTVRNVVDFGAFVDIGVHEDG  
 Sbjct: 601 DIIADLLKPGRDLRDDFEAPILRQDILDLDKLEIGQKLEGTVRNVVDFGAFVDIGVHEDG 660

Query: 688 LIHQSRLIKRRDKKTRKMPLOHPSKYLSVGDIIVTVWVEVDAERSRIGLSLIK 743  
 LIH S + K + HPS+ +SVGD+VTWVW ++D +R ++ LSL+ P  
 Sbjct: 661 LIHISEMSKTF-----VNHPQVSVSGDLVTWVWSKIDLRHKVNLSSLPP 706

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 348

A DNA sequence (GBSx0379) was identified in *S.agalactiae* <SEQ ID 1129> which encodes the amino acid sequence <SEQ ID 1130>. This protein is predicted to be N5,N10-methylenetetrahydromethanopterin reductase homolog. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4864(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

-441-

>GP:AAB94650 GB:U96107 N5,N10-methylenetetrahydromethanopterin  
 reductase homolog [Staphylococcus carnosus]

Identities = 164/300 (54%), Positives = 217/300 (71%), Gaps = 1/300 (0%)

5 Query: 45 VYGIGEHHRDFAVSAPEIVLAAGAVRTNNIRLSSAVTILSSNDPIRVYQQFSTIDALSN 104  
 +YG+GEHHR D+AVS P VLAA A T I+LSSAVT+LSS+DP+ VY++F+T+DA+SN  
 Sbjct: 1 MYGLGEHHRSDYAVSDPVTVLAAAASLTQRIKLSSAVTVLSSDDPVCVYERFATLDAVSN 60

10 Query: 105 GRAEIMAGRGSFIESFPLFGYDLADYDDLDFNEKMDMLLAINSATNLDWKGHILTQTVNERP 164  
 GRAEIM GRGSFIESFPLFGYDL DYD LF EK+++L IN + W+G + +  
 Sbjct: 61 GRAEIMVGRGSFIESFPLFGYDLDDYDRLFVEKLELLKEINQHEVVTWEGTMRPAIKGLG 120

15 Query: 165 IYPRALQRQLPIWVATGGNVDSTIRIAEQGLPIVYATIGGNPKAFRQLVHIYKEVGSRRNG 224  
 +YPRA+Q ++PIW+ATGG +S+IR AE GLPI YA IGGNPK F++ + IY+ V G  
 Sbjct: 121 VYPRAVQDEIPIWLATGGTPESSIRAAEFLPITYAIIGGNPKRFRKNIAIYRAVAESRG 180

20 Query: 225 HKPEQLKVAHWSWGWIEEDNQTAIDRYFFPTKQTVDNIAKGRPHWSEMTKEQYLRSVGPE 284  
 + + VA HSWG+I + ++ A ++ PTK + IAK R +W T+ + R + E  
 Sbjct: 181 YDLADMPVAVHWSWGYIADTDEQAQREFYEPTKVHHEIIAKER-NWPPYTEAHFQREISDE 239

25 Query: 285 GAIFVGSPEVVAHKIIGLVEALELDRFMLHLPVGSMPHKDVLNAIKLYGKEVAPIVRKYF 344  
 GA+FVGSPE VA K+I ++E L L+RFMLH+PVGSMMPH+ ++ AIKLYGK V PI+ YF  
 Sbjct: 240 GAMFVGSPETVARKMIKVIEELGLNRFMLHIPVGSMPHERIMKAIKLYGKRVKPIIEDYF 299

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 349

30 A DNA sequence (GBSx0380) was identified in *S.agalactiae* <SEQ ID 1131> which encodes the amino acid sequence <SEQ ID 1132>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.1310(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 A related GBS nucleic acid sequence <SEQ ID 9715> which encodes amino acid sequence <SEQ ID 9716> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1133> which encodes the amino acid sequence <SEQ ID 1134>. Analysis of this protein sequence reveals the following:

Possible site: 25

45 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.0915(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 20/40 (50%), Positives = 27/40 (67%), Gaps = 3/40 (7%)

55 Query: 4 MAITHKRQDDLESFASFAKVP---KPKKVDSDSKPEQKD 40  
 MAITHK+ D+LE M A FA +P KP +V++D K K+

-442-

Sbjct: 1 MAITHKKNDELEKMLAGFASIPSFDPKPLEVNTDGKLATKE 40

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 5 Example 350

A DNA sequence (GBSx0381) was identified in *S.agalactiae* <SEQ ID 1135> which encodes the amino acid sequence <SEQ ID 1136>. Analysis of this protein sequence reveals the following:

Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1453(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 20 Example 351

A DNA sequence (GBSx0382) was identified in *S.agalactiae* <SEQ ID 1137> which encodes the amino acid sequence <SEQ ID 1138>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -11.15 Transmembrane 216 - 232 ( 210 - 240)

INTEGRAL Likelihood = -9.18 Transmembrane 15 - 31 ( 10 - 39)

INTEGRAL Likelihood = -9.02 Transmembrane 283 - 299 ( 276 - 299)

INTEGRAL Likelihood = -8.76 Transmembrane 128 - 144 ( 119 - 150)

INTEGRAL Likelihood = -4.62 Transmembrane 243 - 259 ( 237 - 265)

INTEGRAL Likelihood = -2.44 Transmembrane 65 - 81 ( 65 - 81)

INTEGRAL Likelihood = -2.44 Transmembrane 94 - 110 ( 93 - 111)

----- Final Results -----

bacterial membrane --- Certainty=0.5458(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12119 GB:Z99105 ycgR [Bacillus subtilis]

Identities = 141/283 (49%), Positives = 198/283 (69%), Gaps = 3/283 (1%)

Query: 10 SVLQWFIFAISIIIEALPFVLLGTILSGIIEVFITPDIVNKFLPKNKLRLVLFGTFFVGFV 69

S LQ +IFISI+IEA+PF+L+G ILSGII++F++ +++ + +PKN+FL VLFG G +

Sbjct: 6 SFLQLNSIFISILIEAIPFILIGVILSGIIMQFVSEEMIRIMPKNRFLAVLFGALAGVL 65

Query: 70 FPSCECGIIPINRFLKVKPSYTAVPFLATAPIINPIVLFATYSAFGNSIRFLILRFVG 129

FP+CECGIIP R L K VP + V F+ TAPIINPIVLF+TY AFGN + R

Sbjct: 66 FPACECGIIPITRLLKGVPLHAGVAFMLTAPIINPIVLFSTYIAFGNRWSVVFYRGGL 125

Query: 130 ATIVAIALGVMLAFLVDDNLIKEDAKPTHFDYSDKKWYQKIFLALAHAIDEFDGTGRYL 189

A V++ +GV+L++ DN L + +P H H + QK+ L HAIDFF G+YL

Sbjct: 126 ALAVSLIIGVILSYQFKDNQLKPDPEGHHHHHGTLL-LQKLGGTLRHAIDFFSVGKYL 184

Query: 190 VFGLTIASAMQIYLPTRVLTIGHSPITAILVMMLLAFILSLCSEADAFIGASLLSTFGI 249

-443-

+ G IA+AMQ Y+ T L IG + +++ LVMM LAF+LSLCSE DAFI +S STF +  
 Sbjct: 185 IIGAFIAAAMQTYVKTSTLLAIGQNDVSSSLVMMGLAFVLSLCSEVDAFIASSFSSTFSL 244

Query: 250 APVMAFLLIGPMIDIKNLMMVNSFKTRFIVQFISVSSLIIII 292

++AFL+ G M+DIKNL+MM+ +FK RF+ F+ ++ +++I+

Sbjct: 245 GSIIAFLVFGAMVDIKNLLMMLAFAFKRFV--FLLITYIVIV 285

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1139> which encodes the amino acid sequence <SEQ ID 1140>. Analysis of this protein sequence reveals the following:

Possible site: 25  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -9.92 Transmembrane 216 - 232 ( 211 - 237)  
 INTEGRAL Likelihood = -9.45 Transmembrane 283 - 299 ( 276 - 299)  
 INTEGRAL Likelihood = -8.76 Transmembrane 128 - 144 ( 119 - 150)  
 INTEGRAL Likelihood = -7.80 Transmembrane 15 - 31 ( 10 - 39)  
 INTEGRAL Likelihood = -5.47 Transmembrane 243 - 259 ( 237 - 265)  
 INTEGRAL Likelihood = -2.44 Transmembrane 65 - 81 ( 65 - 81)  
 INTEGRAL Likelihood = -2.44 Transmembrane 94 - 110 ( 93 - 111)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.4970(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB12119 GB:Z99105 ycgR [Bacillus subtilis]  
 Identities = 143/288 (49%), Positives = 196/288 (67%), Gaps = 1/288 (0%)  
 Query: 10 SVLQWFQFAIFMSIIIEALPFVLLGTILSGCIEVFVTPPELVQKYLPKQKCLRILFGTFVGVFV 69  
 S LQ +IF+SI+IEA+PF+L+G ILSG I++FV+ E++ + +PK + L +LFG G +  
 Sbjct: 6 SFLQLNSIFISILIEAIPFILIGVILSGIIQMFVSEEMIRIMPKNRFLAVLFGALAGVL 65  
 Query: 70 FPSCECGIPIINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNSLRFLILRLVG 129  
 FP+CECGIPI R L K VP + V F+ TAPIINPIVLF+TY AFGN + R  
 Sbjct: 66 FPACCEGIIPITRLLKGVPLHAGVAFMLTAPIINPIVLFSTYIAFGNRWSVVFYRGGL 125  
 Query: 130 AALVAITLGVMLAFIVDDNILKDNAQPVHFHDYSHESLPKRIYLALVHAIDFFDTGRYL 189  
 A V++ +GV+L++ DN L +P H H + H +L +++ L HAIDFF G+YL  
 Sbjct: 126 ALAVSLIIGVILSYQFKDNQLKPDPEGH-HHHHGTLLQLKGGTLRHAIDFFSVGKYL 184  
 Query: 190 VFGTLIASAMQIYVPTRVLT'TIGHNPLITAILMMLMAFILSLCSEADAFIGASLLSTFGV 249  
 + G IA+AMQ YV T L IG N +++ L+MM +AF+LSLCSE DAFI +S STF +  
 Sbjct: 185 IIGAFIAAAMQTYVKTSTLLAIGQNDVSSSLVMMGLAFVLSLCSEVDAFIASSFSSTFSL 244  
 Query: 250 APVLAFLLLIGPMVDIKNLMMVKAFKGRFIVQFIGSVLMIYVYCLLV 297  
 ++AFL+ G MVDIKNL+MM+ AFK RF+ I V+++ LLV  
 Sbjct: 245 GSIIAFLVFGAMVDIKNLLMMLAFAFKRFVFLITYIVIVVLGSLIV 292

An alignment of the GAS and GBS proteins is shown below:

Identities = 248/300 (82%), Positives = 278/300 (92%)  
 Query: 1 MDIFNQLPDSVLQWFAIFISIIIEALPFVLLGTILSGIIEVFITPDIVNKFLPKNKFLRV 60  
 M +F+ LP SVLQWFAIF+SIIEALPFVLLGTILSG IEVF+TP++V K+LPK K LR+  
 Sbjct: 1 MSIFSNNLPPSVLQWFAIFMSIIIEALPFVLLGTILSGCIEVFVTPPELVQKYLPKQKCLR 60  
 Query: 61 LFGTFVGVFVFPSCCEGIPIINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNSI 120  
 LFGTFVGVFVFPSCCEGIPIINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNS+  
 Sbjct: 61 LFGTFVGVFVFPSCCEGIPIINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNSL 120  
 Query: 121 RFLILRFVGATIVAIALGVMLAFIVDDNILKEDAKPTHFHDYSDKKWYQKIFLALAHAI 180  
 RFLILR VGA +VAI LGVMLAF+VDDNILK++A+P HFHDYS + ++I+LAL HAID  
 Sbjct: 121 RFLILRLVGAALVAITLGVMLAFIVDDNILKDNAQPVHFHDYSHESLPKRIYLALVHAID 180  
 Query: 181 EFFDTGRYLVFGTLIASAMQIYLPTRVLT'TIGHSPITAILVMMLLAFILSLCSEADAFI 240

-444-

EFFDTGRYLVFGTLIASAMQIY+PTRVLT TIGH+P+TAIL+MML+AFILSLCSEADAFIG  
 Sbjct: 181 EFFDTGRYLVFGTLIASAMQIYVPTRVLT TIGHNPLTAILMMLMAFILSLCSEADAFIG 240

Query: 241 ASLLSTFGIAPVMAFLLIGPMIDIKNLMMMVNSFKTRFIVQFISVSSLIIIIYCLFVGVI 300  
 ASLLSTFG+APV+AFLLIGPM+DIKNLMMMV +FK RFIVQFI VS L+I +YCL VGV+

Sbjct: 241 ASLLSTFGVAPVLAFLLLIGPMVDIKNLMMMVKA FKGRFIVQFIGVSVLMIAVYCLLVGVL 300

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 352

A DNA sequence (GBSx0383) was identified in *S.agalactiae* <SEQ ID 1141> which encodes the amino acid sequence <SEQ ID 1142>. Analysis of this protein sequence reveals the following:

Possible site: 13  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.4703(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 353

A DNA sequence (GBSx0384) was identified in *S.agalactiae* <SEQ ID 1143> which encodes the amino acid sequence <SEQ ID 1144>. Analysis of this protein sequence reveals the following:

Possible site: 50  
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -8.44	Transmembrane	45 - 61 ( 39 - 65)
INTEGRAL	Likelihood = -8.12	Transmembrane	83 - 99 ( 77 - 101)
INTEGRAL	Likelihood = -0.00	Transmembrane	2 - 18 ( 1 - 19)

----- Final Results -----

bacterial membrane	---	Certainty=0.4376(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 8559> which encodes amino acid sequence <SEQ ID 8560> was also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1    Crend: 2  
 SRCFLG: 0  
 McG: Length of UR: 8  
 Peak Value of UR: 2.23  
 Net Charge of CR: 1  
 McG: Discrim Score: 0.46  
 GvH: Signal Score (-7.5): -3.54  
 Possible site: 42  
 >>> Seems to have an uncleavable N-term signal seq

Amino Acid Composition: calculated from 1

ALOM program    count: 2    value: -8.44    threshold: 0.0

INTEGRAL	Likelihood = -8.44	Transmembrane	37 - 53 ( 31 - 57)
INTEGRAL	Likelihood = -8.12	Transmembrane	75 - 91 ( 69 - 93)
PERIPHERAL	Likelihood = 2.76		200

-445-

modified ALOM score: 2.19  
icml HYPID: 7 CFP: 0.438

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.4376(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12118 GB:Z99105 ycgQ [Bacillus subtilis]  
Identities = 100/290 (34%), Positives = 159/290 (54%), Gaps = 25/290 (8%)

Query: 9 MIRFLILAGYFELSMYKLKSGKLNQYINTHTYLYISMVLSFILAIIVQLIIWVKNMKMH 68  
M R L+L G+ +L SG L +YIN Y YL++I++ L IL VQ +++K+ +  
Sbjct: 1 MFRLLVLMGFTFFFYHLHASGNLTKEYINMKYAYLSFIAIFLLAILTAVQAYLFIKSPEKS 60

Query: 69 SHLHGKIA-----KSTSP-----MILVFPVLVGLLVPTVSLDSTTVSAGYN 110  
H H + P ++ +FP++ G+ P +LDS+ V KG++  
Sbjct: 61 GHHDHDCGCGHDHEHDHEQNKFYQRYLIYVFLFPLVSGIFFPIATLDSIVKTKGFS 120

Query: 111 FPLAAGSTGTVSQDGRVQYLKPDSTSTYFTSSAYEKEMQKELKKYKSGTTLTITTENYME 170  
F A S SQ QYL+PD S Y+ +Y+K+M++ KY +++T +++++  
Sbjct: 121 FK-AMESGDHYSQ----TQYLRPDASLYYAQDSYDKQMKQLFNKYSSKKEISLTDDDFLK 175

Query: 171 VMELIYLYPEQFMDRQIQYTGfVY-NEPKHEGYQFIFRFGIIHICIADSGVYGLLT-GNQ 228  
ME IY YP +F+ R I++ GF Y ++ F+ RFGIIHICIADSGVYG+L  
Sbjct: 176 GMETIYNYPGEFLGRTIEFHGFAYKGNAINKNQLFVLRFGLIHCIIADSGVYGMLEFPKD 235

Query: 229 KSYPDNTWVTVRGTIKSEYNQLLQONLFPVLHIEESRQVSKANNPYVYRVF 278  
D+ W+ ++GT+ SEY Q + LPV+ + + K ++PYVYR F  
Sbjct: 236 MDIKDDEWIHIKGTLA SEYYQPFKSTLPVVKVTDWNTIKKPDPPYVYRGF 285

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1145> which encodes the amino acid sequence <SEQ ID 1146>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -8.33 Transmembrane 83 - 99 ( 74 - 101)  
INTEGRAL Likelihood = -6.21 Transmembrane 42 - 58 ( 39 - 62)

----- Final Results -----

bacterial membrane --- Certainty=0.4333(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9115> which encodes the amino acid sequence <SEQ ID 9116>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 54

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -8.33 Transmembrane 75 - 91 ( 66 - 93)  
INTEGRAL Likelihood = -6.21 Transmembrane 34 - 50 ( 31 - 54)  
PERIPHERAL Likelihood = 2.76

----- Final Results -----

bacterial membrane --- Certainty= 0.433(Affirmative) < succ>  
bacterial outside --- Certainty= 0.000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 208/279 (74%), Positives = 244/279 (86%), Gaps = 1/279 (0%)

-446-

Query: 1 MFICGGNIMIRFLILAGYFELSMYKLKSGKLNQYINTHTYTYLAYISMVLSFILAIIVQLII 60  
 +F CCG +MIRFLILAGYFEL+MYL+LSGKL+QYIN Y+YLAYISM+LSFILA+VQL  
 Sbjct: 1 LFTCCGALMIRFLILAGYFELTMYLQLSGKLDQYINVRYSYLAYISMILSFILALVQLYT 60

5 Query: 61 WVKNMKMHSHLHGKIAKSTSPMILVFPVLVGLLVPTVSLDSTTVSAKGYNFPLAAGSTGT 120  
 W+KN+K+HSHL GKIA+ TSP ILVFPVL+GLLVPTV+LDSTTVSAKGY FPLAAG++ T  
 Sbjct: 61 WMKNIKVHSHLTGKIARLTSPFILVFPVLIGLLVPTVTLDSSTTVSAKGYTFPLAAGASKT 120

10 Query: 121 -VSQDQTRVQYLKPDSTSTYFTSSAYEKEMQKELKKYKSGTLTITTENYMEVMELIYLYP 179  
 VS DGT +QYLKPDTS YFT SAY+KEM++EL KYKG +TITTENYMEVMELIYLYP  
 Sbjct: 121 GVSDDGTTIQYLKPDTSIYFTKSAYQKEMRQELHKYKGGKPVTTITTENYMEVMELIYLYP 180

15 Query: 180 EQFMDRQIQYTGfVYNEPKHEGYQFIFRFGIIHICIADSGVYGLLTTGNQKSYPDNTWVTV 239  
 ++F+DR IQYTGfVYNEP H+ YQF+FRFGIIHICIADSGVYGLLTTGNQ SYP+NTW+TV  
 Sbjct: 181 DEFLDRDIQYTGfVYNEPGHDNYQFLFRFGIIHICIADSGVYGLLTTGNQTSYPNNTWLTW 240

20 Query: 240 RGTIKSEYNQLLQONLPVLHIEESRQVSKANNPYVYRVF 278  
 +G + EY++ L+Q+LPVL + E Q + NNPYVYRVF  
 Sbjct: 241 KGRHMEYDKNLEQHLPLVQLAEVHQTKPEPNPYVYRVF 279

SEQ ID 8560 (GBS235d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 14 & 15; MW 48.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 17 & 18; MW 23.4kDa), in Figure 150 (lane 15; MW 23kDa) and in Figure 182 (lane 5; MW 23kDa).

GBS235d-His was purified as shown in Figure 235, lane 6-7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 354

A DNA sequence (GBSx0385) was identified in *S.agalactiae* <SEQ ID 1147> which encodes the amino acid sequence <SEQ ID 1148>. This protein is predicted to be signal recognition particle (ftsY). Analysis of this protein sequence reveals the following:

Possible site: 57  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3301(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB06205 GB:AP001515 signal recognition particle (docking protein) [Bacillus halodurans]  
 Identities = 175/304 (57%), Positives = 227/304 (74%)

Query: 233 EKYNRSLKKTRTGFSARLNAFLSNFRRVDEEFFEELIEMILSDVGVNVATQLTEDLRYE 292  
 EK+ L+KTR F+ ++N + +R VDE+FFEELEEL+LI +DVGVL L E+L+ E  
 Sbjct: 20 EKFKAGLEKTRDSFAGKMNDLVYKYRSVDEDFEELIILIGADVGVTTVMIDLVEELKDE 79

Query: 293 AKLENAKKSEDLRVIVEKLVETIYEKDIYNEAINFQEGTLVMLFVGVNGVGKTTSIGKL 352  
 + +N K S+D++ +I EKL E+ EK+G E GL+V+L VGVNGVGKTTSIGKL  
 Sbjct: 80 VRRQNIKDSKDIQPIISEKLAELLEKEGGETEVNLQAGLSVILVGVNGVGKTTSIGKL 139

Query: 353 AHQYKSQGGKVMVAADTFRAGAVAQLVEWGRRVDVPVVTGEEKADPASVVFDMGMEKAVA 412  
 AH YK QGKKV+L A DTFRAGA+ QL WG R V V+ E +DPA+V+FD ++ A +  
 Sbjct: 140 AHMYKQGGKVVILAAGDTFRAGAIEQLEWGERAGVDVIKQSEGSDPAVMFDAIQAAS 199

Query: 413 QGVDVLLIDTAGRLQNKENLMAELEKIGRIIKRVVPDAPHETLLALDASTGQNALSQAKE 472



-447-

+ D+L+ DTAGRLQNK NLM ELEK+ R+I R +P APHE L+ALDA+TGQNA+SQAK  
 Sbjct: 200 READILICDTAGRLQNKVNLMELEKVKRVISREIPGAPHEVLIALDATTGQNAMSQAKT 259

Query: 473 FSKITPLTGLILTKIDGTAKGGVVLAIQELDIPVKFIGFGEKIDDIGEFNSED FMRGLL 532  
 F + T +TG+ILTK+DGTAKGG+VLAI R ELDIPVKF+G GEKIDD+ F+SE F+ GL  
 Sbjct: 260 FKETTDVTGIILTKLDGTAKGGIVLAIRHELDIPVKFVGLGEKIDDLQPFDSQFVYGLF 319

Query: 533 EGIL 536  
 + ++  
 Sbjct: 320 KDMV 323

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1149> which encodes the amino acid sequence <SEQ ID 1150>. Analysis of this protein sequence reveals the following:

Possible site: 60  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4384(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 339/549 (61%), Positives = 404/549 (72%), Gaps = 46/549 (8%)

Query: 1 MGLFDRLFGHKKKDKEPEIEASESVVLEDEDSVIDKEEGSNFSKESTLNRTSEVPVAEDD 60  
 MGLFDRLFG K+ K E + E+++ E KEE S + E ++ + +  
 Sbjct: 1 MGLFDRLFGKKETPKVAEEKLEENLLTE----TTQKEELSEKANEQ-----DKIEAVQQE 51

Query: 61 SFLELERDTALSESHPVTSEIHPLESEDTEIPVKEDDSFLELEDRAKTKVADTSEVEN 120  
 ++ + A S + P + ++ L E+T D + DT+E  
 Sbjct: 52 ---DVSSEGAGSVENGPEAASVNALVEEETG-----DNSNHPSEDTNEF-- 92

Query: 121 VVPDSTTLSDNVSASEASFSQSDQFSETPLQSEMS--SGKTEVQTESED 178  
 D T L VS S+++ S+ + L D +QF Q + S S E S++  
 Sbjct: 93 -AADKTDLK--VSELSQSTASEPKDLVDQPVVEQFPPTKQQAQASNDSEAVDTSKEQ 149

Query: 179 SAADAFLADYYAKRKAIEKEISSNSLST-----DESEFSEAQEVLSQSQ--DTIK 227  
 S++ + DYY ++ A+EK + + +T E++ S + E SQ++A DTI  
 Sbjct: 150 SSSQQVMEDYYRKAALKESLQEKAAATVPVMPPEEVPQENQASTSAEA-SQNKATHDTIP 208

Query: 228 AESQEEKYNRSLKKTRTGFSARLNFLSNFRRVDEEFFEELMILSDVGVNVATQLTE 287  
 E+ +EKY RSLKKTRTGFSARLN+F +NFRRVDEEFFE+LEMLILSDVGV+VAT LTE  
 Sbjct: 209 -ETDQEKYKRLKKTRTGFSARLNSFFANFRRVDEEFFEDLEMLILSDVGVHVATTLTE 267

Query: 288 DLRYEAKLENAKKSEDLKRVIVEKLVEIYEKDG YNEAINFQEGTLVMLFVGNGVGKTT 347  
 +LRYEAKLENAKK + LKRVIVEKLV+IYEKDG YNEAIN+Q+GLTVMLFVGNGVGKTT  
 Sbjct: 268 ELRYEAKLENAKKPDALKRVIVEKLVDIYEKDGRYNEAINYQDGLTVMLFVGNGVGKTT 327

Query: 348 SIGKLAHQYKSQGGKVMLVAADTFRAGAVAQLVEWGRRVDVPVVTGEEKADPASVVFDDGM 407  
 SIGKLA++YK +GKKVMLVAADTFRAGAVAQLVEWGRRVDVPV+TG EKADPASVVFDDGM  
 Sbjct: 328 SIGKLAYRYKQEGKKVMLVAADTFRAGAVAQLVEWGRRVDVPVITGPEKADPASVVFDDGM 387

Query: 408 EKAVAQGVDDVLLIDTAGRLQNKENLMAELEKIGRIIKRVVPDAPHETLLALDASTGQNAL 467  
 EKAVA+GVD+LLIDTAGRLQNKENLMAELEK+GRIIKRV+PDAPHETLLALDASTGQNAL  
 Sbjct: 388 EKAVAGVDILLIDTAGRLQNKENLMAELEKMGRIIKRVLPDAPHETLLALDASTGQNAL 447

Query: 468 SQAKEFSKITPLTGLILTKIDGTAKGGVVLAIQELDIPVKFIGFGEKIDDIGEFNSED 527  
 SQAKEFSKITPLTGLILTKIDGTAKGGVVLAIQELDIPVKFIGFGEK+DDIGEF+SEDF  
 Sbjct: 448 SQAKEFSKITPLTGLILTKIDGTAKGGVVLAIQELDIPVKFIGFGEKVDDIGEFHSEDF 507

Query: 528 MRGLLEGIL 536  
 M+GLLEGIL  
 Sbjct: 508 MKGLLEGIL 516

-448-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 355

A DNA sequence (GBSx0386) was identified in *S.agalactiae* <SEQ ID 1151> which encodes the amino acid sequence <SEQ ID 1152>. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10           bacterial cytoplasm --- Certainty=0.3592(Affirmative) < succ>  
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15       >GP:AAA62048 GB:L10328 f270 [Escherichia coli]  
         Identities = 101/273 (36%), Positives = 160/273 (57%), Gaps = 10/273 (3%)

Query: 4    IKILALDLDGTLFTTDDKKVSEENKVALKAAREKGIKVVITTTGRPLKAIGNLLEDLELVSD 63  
          IK++A+D+DGTL   D   +S   K A+ AAR +G+ VV+TTGRP   + N L++L +  
20   Sbjct: 3   IKLIAIDMDGTLTLLPDHTTISPAVKNAIAAARARGVNVVLTTRGPYAGVHNYLKELHMEQP 62

Query: 64   EDYSITFNGGLVQQNT-GKILAKTAMTRQEVEDIHEELYQVGLPTDILSEGTIVYS----I 118  
          DY IT+NG LVQ+   G   +A+TA++   +   +   +   +VG   L   T+Y+   I  
25   Sbjct: 63   GDYCITYNGALVQKAADGSTVAQTALSYDDYRXLEKLSREVGSFHALDRTTLYTANRDI 122

Query: 119   ANKGHHSQYHLANPLLEFIEVDDLEQVPKDVVYNKIVSVIDATYLDQQIAKLPDRCLKVDY 178  
          +   H   +   PL+ F E   E++   +   + K++   +   +   LDQ IA++P   +K Y  
30   Sbjct: 123   SYYT VHESFVATIPLV-FCEA---EKMDPNTQFLKVMIMIDEPAILDQAIARIPQXVKEY 178

Query: 179   EMFKSRDIILELMPKGVHKAVGLELTLTKHLGLDSSQVMAMGDEANDLSMLEWAGLGVAMA 238  
          + KS   LE++ K V+K G++ L   LG+   ++MA+GD+ ND++M+E+AG+GVAM  
35   Sbjct: 179   TVLKSAPYFLEILDKRVNKGTVGKSLADVLGIKPEEIMAGDQENDIAMIEYAGVGVAMD 238

Query: 239   NGIPEAKAIAKATTICNNDESGVAEAIKGYILS 271  
          N IP K +A   T   +N E GVA AI KY+L+  
40   Sbjct: 239   NAIPSVKEVANFVT-KSNLEDGVAFAIEKYVLN 270

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1153> which encodes the amino acid sequence <SEQ ID 1154>. Analysis of this protein sequence reveals the following:

40       Possible site: 32  
         >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45           bacterial cytoplasm --- Certainty=0.3502(Affirmative) < succ>  
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 180/273 (65%), Positives = 218/273 (78%), Gaps = 1/273 (0%)

50       Query: 3    DIKILALDLDGTLFTTDDKKVSEENKVALKAAREKGIKVVITTTGRPLKAIGNLLEDLELVSD 62  
          +I+ILALDLDGTL+ T+K V++ NK AL AAREKG+KVVITTTGRPLKAIGNLLE+L+L+  
          Sbjct: 2   NIRILALDLDGTLNTEKIVTDANKALAAAREKGVKVVITTTGRPLKAIGNLLEELDLLD 61

55       Query: 63   DEDYSITFNGGLVQQNTGKILAKTAMTRQEVEDIHEELYQVGLPTDILSEGTIVYSIANK- 121  
          +DYSITFNGGLVQ+NTG++L K++++ +V I + L VGLPTDI+S G VYSI +K  
          Sbjct: 62   HDDYSITFNGGLVQRNTGEVLDKSSLSFDQVCQIQQALEAVGLPTDIISGGDVYSIPS 121

Query: 122   GHHSQYHLANPLLEFIEVDDLEQVPKDVVYNKIVSVIDATYLDQQIAKLPDRCLKVDYEMF 181

-449-

G HSQVHLANPLL FIEV + ++PKD+ YNKIV+V D +LDQQI KL L D+E F  
 Sbjct: 122 GRHSQVHLANPLLTFIEVTSVAELPKDITYNKIVTVTDPDFLDQQI IKLSPSLFDFFAF 181

Query: 182 KSRDIIELMPKGVHKAVGLELLTKHLGLDSSQVMAMGDEANDLSMLEWAGLGVMANGI 241  
 KSRDII E+MPKG+ KA GL LL +HLGLD+ VMAMGDEAND +MLEWAGLGVMANG+  
 Sbjct: 182 KSRDIIIFEIMPKGIDKAFGLNLLCQHLGLDARHVMAMGDEANDFAMLEWAGLGVMANGV 241

Query: 242 PEAKAIAKATTICNNDESGVAEAIGKYILSEEN 274  
 AKA A A T NDESGVAEA+ +IL EE+  
 Sbjct: 242 SGAKADADAVTTLTNDDESGVAEAVKTFILEEES 274

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 356

- 15 A DNA sequence (GBSx0387) was identified in *S.agalactiae* <SEQ ID 1155> which encodes the amino acid sequence <SEQ ID 1156>. Analysis of this protein sequence reveals the following:

Possible site: 49  
 >>> Seems to have no N-terminal signal sequence

- 20 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.4648(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 25 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA35556 GB:D90723 Hypothetical 30.2 kd protein in idh-deoR  
           intergenic region. [Escherichia coli]  
   Identities = 91/264 (34%), Positives = 146/264 (54%), Gaps = 4/264 (1%)

- 30 Query: 2 IKLVATDMDGTFLDENGTYDKRLANVLKKFKEQGIVFTAASGRSLLSLEQLFADFRDQM 61  
           IKL+A DMDGTFL + TY+++R ++ K QGI F ASG L F + +++  
   Sbjct: 4 IKLIAVDMMDGTFLSDQKTYNRERFMAQYQMQKAQGIRFVVASGNQYYQLISFFPEIANEI 63
- 35 Query: 62 AFIAENGSAAVLFNRLAYEQHLSREQYLDIIDHLSKSPYMENNEYVLSGKDGAYILSDAN 121  
           AF+AENG V + + LS++ + +++HL P + E + GK+ AY L +  
   Sbjct: 64 AFVAENGWVSEKDVFNELSKDAFATVVEHLLTRPEV---EIIACGKNSAYTLKKYD 120
- 40 Query: 122 PDYIEFITHYYDNLQKVSHFEDVDDIIFKVTANFTEETVRQAEWVNQAI-PYATAVTTG 180  
           YY L+ V +F++++DI FK N ++E + Q ++ +++AI +V TG  
   Sbjct: 121 DAMKTVAEMYHRLLEYVDNFDNLEIDIFFKFGNLNLSDELIPQVQKALHEAIGDIMVSVHTG 180
- 45 Query: 181 FKSIDIILSSVKNRNGLEHLCEQYGIRAEVLSFGDNINDLEMLEWSGKAIATENARPEV 240  
           SID+I+ V+K NGL L + +GI EV+ FGD ND+EML +G + A ENA V  
   Sbjct: 181 NGSIDLIIPGVHKANGLRQLQKLWGIDDSEVVVFGDGGNDIEMLRQAGFSFAMENAGSAV 240
- Query: 241 KEIADCIIGHHNNQAVMAYLESMV 264  
           A G +N + V+ ++ ++  
   Sbjct: 241 VAAAKYRAGSNNREGVLDVIDKVL 264

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1157> which encodes the amino acid sequence <SEQ ID 1158>. Analysis of this protein sequence reveals the following:

Possible site: 49  
 >>> Seems to have no N-terminal signal sequence

- 55 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.3401(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 60 An alignment of the GAS and GBS proteins is shown below:

-450-

Identities = 138/265 (52%), Positives = 193/265 (72%), Gaps = 1/265 (0%)

Query: 1 MIKLVATDMDGTFLDENGTYDKKRLANVLKKFKEQGIIVFTAAASGRSLLSLEQLFADFRDQ 60  
 Sbjet: 1 MIKL+ATDMDGTFL E+GTY++++LA +L K E+GI+F +SGRSLL+++QLF F DQ

Query: 61 MAFIAENGSAAVLFNRLAYEQHLSREQVLDIIDHLSKSPYEMENNEYVLSGKDGAYILSDA 120  
 +A IAENGs + + +++EQY ++ + +P+ V SG+ AYIL A  
 Sbjet: 61 IAVIAENGSVVQYRGEILFADMMTKEQYTEVAKKILANPHYVETGMVFSGQKAAYILKGA 120

Query: 121 NPDYIEFITHYYDNLQKVSHFEDVD-DIIFKVTANFTEETVRQAEWVNQAIPYATAVTT 179  
 + +YI+ HYY N++ ++ FED++ D IFKV+ NFT TV + +W+NQA+PYATAVTT  
 Sbjet: 121 SEEYIQTKHYIANVKVINGFEDMENDAIFKVSTNFTGHTVLEGSDDLNLQALPYATAVTT 180

Query: 180 GFKSIDIIILSSVKNRNGLEHLCEQYGIIRAEVLSFGDNINDLEMLWSGKAIATENARPE 239  
 GF SIDIIL VNK G+EHLC+ GI+ E ++FGDN ND +MLE++G+AIATENARPE  
 Sbjet: 181 GFDSIDIILKEVNKGFGMEHLCQALGIKKAETIAFGDNFNQYQMLEFAGRAIATENARPE 240

Query: 240 VKEIADCIIGHNNQAVMAYLESMV 264  
 +K I+D +IGH N+ AV+ YL+ +V  
 Sbjet: 241 IKVISDQVIGHCNDGAVLTYLKGGLV 265

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 25 Example 357

A DNA sequence (GBSx0388) was identified in *S.galactiae* <SEQ ID 1159> which encodes the amino acid sequence <SEQ ID 1160>. Analysis of this protein sequence reveals the following:

Possible site: 18  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2428(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 40 Example 358

A DNA sequence (GBSx0389) was identified in *S.galactiae* <SEQ ID 1161> which encodes the amino acid sequence <SEQ ID 1162>. This protein is predicted to be p115 protein (smc). Analysis of this protein sequence reveals the following:

Possible site: 55  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -4.99 Transmembrane 1092 -1108 (1088 -1110)

----- Final Results -----  
 bacterial membrane --- Certainty=0.2996(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9713> which encodes amino acid sequence <SEQ ID 9714> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB13467 GB:Z99112 chromosome segregation SMC protein homolg  
[Bacillus subtilis]  
Identities = 458/1193 (38%), Positives = 728/1193 (60%), Gaps = 27/1193 (2%)

5 Query: 1 MFLKEIEMQGFKSFADKTKVEFDQGVTA VVGPNNGSGKSNITESLRWALGESSAKSLRGGK 60  
MFLK +++ GPKSFA++ V+F +GVTAVVGPNNGSGKSNIT+++RW LGE SA+SLRGGK  
Sbjct: 1 MFLKRLDVGFKSFAERISVDFVKGVTA VVGPNNGSGKSNITDAIRWVLGEQSARSLRGGK 60

10 Query: 61 MPDVIFAGTENRKPLNYA QVSVTLDNSDHF IENIAD E V R VERRIFRNGDSEY LIDGRKVR 120  
M D+IFAG+++RK LN A+V++TLDN DHF+ EV V RR++R+G+SE+LI+ + R  
Sbjct: 61 MEDII FAGSDSRKRLNLA E V T L T L D N D D H F L P I D F H E V S V T R R V R S G E S E F L I N N Q P C R 120

15 Query: 121 LRDIHDLFMDTGLGRDSFSIISQGRVEAIFNSKPEERRAIFEEAAGVLKYKTRKKTQSK 180  
L+DI DLFMD+GLG+++FSIISQG+VE I +SK E+RR+IFEEAAGVLKYKTRKK+ ++K  
Sbjct: 121 LKDIIDL FMD SGLGKEAFS IISQ G K V E E I L S S K A E D R R S I F E E A A G V L K Y K T R K K K A E N K 180

20 Query: 181 LEQTQGNLDRLEDIIYELDMQVQPLEKQASIAKRFLVLDEERQGLHLSILIEDILQHQS 240  
L +TQ NL+R+EDI++EL+ QV+PL+ QASIAK +L +E + + + + + DI +  
Sbjct: 181 LFETQDNLNRVEDILHELEGQVEPLKIQASIAKDYLEKKKELEHVEIALTAYDIEKLHGK 240

25 Query: 241 LTTVEEKLITVRKELATYYQQRQSLDENQSLKQKRHHLSEEIEAKQLALLDVTPLKSDL 300  
+T++EK+ ++E + E + + + K L E + Q LL ++ L  
Sbjct: 241 WSTLKEKVQMAKEEELAESSAISAKEAKIEDTRDKIQALDES VNELQQVLLVTSEELEKL 300

30 Query: 301 ERQIDLIRLESNQKAEKKEEAGORLAELEIKAKDCSDQITQKNIETTLSEKIAQIRSEI 360  
E + + + + + + +E+ + + + + K + + + + + TL ++ Q+R+++  
Sbjct: 301 EGRKEVLKERKKNVQNGEQLEEAIVQFQQKETVLKEELSKQEAVFETLQAEVKQLRAQV 360

35 Query: 361 VSTESSLERFSTNPDIIEKLREDFVTLMQEEADTSNALTALLADIENQKQASQAKSQEI 420  
+ +L + N ++ IE+L+ D+ L+ +A N L LL D +Q + + +  
Sbjct: 361 KEKQQALSLHNENVEEKIEQLKSDYFELLNSQASTRNEL-QLLDDQMSQSAVTLQRLADN 419

40 Query: 421 QEVSKNLEVLKSNKVALE-RFEAAKKNVRQLLSHYQDLGQTLQNLGEYKKNQSSILFDH 479  
E S K A E F ++ + + Y+D+ + + +Y+ +S L+  
Sbjct: 420 NEKHLQERHDISARKAACETEFARIEQEIHSQVGAYRDMQTKYEKKRQYKKNESALYQA 479

45 Query: 480 LDEIKSKQARISSLESILKNHSNFYAGVKSVLQAKDQLGGIIGAVSEHLSFDKHYQTALE 539  
++ +++ LE++ + S FY GVK VL+AK++LGGI GAV E +S ++ Y+TA+E  
Sbjct: 480 YQYVQQARSKDMLETMQGDGSGFYQGVKEVLKAKERLGGIRGAVLELISLEQKYETAIE 539

50 Query: 540 IALGGSSQHIIVEDESAAKRSIAFLKKNRQGRATFLPLTTIKPRELAQHYSKLQSSQGF 599  
IALG S+QH++ +DE +A+++I +LK+N GRATFLPL+ I+ R+L F  
Sbjct: 540 IALGASAQHVVTDDQSARKAIQYLKQNSFGRATFLPLSVIRDRQLQSRDAETAARHSSF 599

55 Query: 600 LGIASLVTYDQRLSNIFKNNLGLTAIFDVTVDNANVAARQLNYQVRLVTLDTGTELRPGGS 659  
LG+ASELVT+D ++ +N LG I + + AN A+ L ++ R+VTL+G + PGGS  
Sbjct: 600 LGVASELVTFDPAYRSVIQNLGLTGLTITLKGANELAKLLGHRYRIVTLEGDVVNPGGS 659

60 Query: 660 YSGGANRQNNTVFI--KPELDNLKKELKQAQSKQLIQEKEVATLLEQLKEKQETLAQLKN 717  
+GGA ++ N + EL+++ K L + + K + E+EV TL +++ ++ LA L+  
Sbjct: 660 MTGGA V K K N N S L G R S R E L E D V T K R L A E M E E K T A L L E Q E V K T L K H S I Q D M E K K L A D L R E 719

65 Query: 718 DGEQARLEEQRADIEYQQLSEKLADLNKLYNGLQLSSGALEQTTSENE--KNRLEKELEQ 775  
GE RL++Q + +L ++N AL ++ E + K +LE+EL  
Sbjct: 720 TGEGLRLKQDVKGQLYELQVAEKNINTHLELYDQEKSALESSEDERKVRKRKLEELSA 779

70 Query: 776 FAIKKEELTTSIAQIKEDKDSIQEKVNNTLLSEAQLEERDLLNEQKFERANCTRL--- 832  
+ K ++L I ++ + K + +L+ L+E ++ K E N RL  
Sbjct: 780 VSEKMKQLEEDIDRLTKQKQTSSTKESLSNELTELKIAAKKEQACKGEEDNLARLKE 839

75 Query: 833 ----EITLSEIKRDISNLQTLLSHQDSQLDKEELPRIEKQLLQVNNRRENDEEKLVS LRF 888  
E+ L E K D+S L + +S S E++L + + ND+ K + L  
Sbjct: 840 LTETELALKEAKEDLSFLTSEMSSSTSG-----EEKLEEA AKHLNDKTKTIELIA 890

80 Query: 889 ELEDCEAALDDLAASLAKEGQKNESLIRQQAQL---ESQCEQLSQQLMIFSRQLSEDYQ 944  
D L + +E ++ + L +Q+ L E + ++ +L + L E+Y

-452-

Sbjct: 891 LRRDQRIKLQHGLDTERELKEMKRLYKQKTTLLKDEEVKLGRMEVELDNLLQYLREEYS 950

Query: 945 MTLDEAKVKANVLEDILMAREQLKSLQAKIKALGPVNIDAIAQFEEVHERLTFNLTRQDD 1004  
 ++ + AK K + D AR+++K ++ I+ LG VN+ +I +FE V+ER FL+ Q++D

5 Sbjct: 951 LSFEGAKEKYQLETDPEEARKRVLKIKLAIEELGTVNLGSIDEFERVNERYKFLSEQKED 1010

Query: 1005 LVHAKNLLLETITDMDEVKTRFKSTFEAIRHSFKETTFVQMFGGGSADLILTE-GDLLSA 1063  
 L AKN L + I +MD+E+ RF TF IR F + F +FGGG A+L LT+ DLL +

10 Sbjct: 1011 LTEAKNTLFQVIEEMDEEMTKRFNDTFVQIRSHFDQVFRSLFGGGAELRLTDPNDLLHS 1070

Query: 1064 GVDISVQPPGKKIQSLNLSMSGGEKALSALALLFAIRVKTIPFVILDEVEAALDEANVKR 1123  
 GV+I QPPGKK+Q+LNL+SGGE+AL+A+ALLF+I++V+ +PF +LDEVEAALDEANV R

15 Sbjct: 1071 GVEIIAQPPGKKLQNLNLLSGGERALTAIALLLFSILKVRPVPFCVLDEVEAALDEANVFR 1130

Query: 1124 FGDYLNRFDKSSQFIVVTHRKGTMASADSIYGVTMQESGVSKIVSVKLKEAQE 1176  
 F YL ++ +QFIV+THRKGTM AD +YGVTMQESGVSK++SVKL+E +E

Sbjct: 1131 FAQYLKKYSSDTQFIVITHRKGTMEEADVLYGVTMQESGVSKIVSVKLEETKE 1183

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1163> which encodes the amino acid  
 20 sequence <SEQ ID 1164>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -4.99 Transmembrane 1092 -1108 (1088 -1110)

----- Final Results -----

bacterial membrane --- Certainty=0.2996(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB13467 GB:Z99112 chromosome segregation SMC protein homolg  
 [Bacillus subtilis]

Identities = 441/1192 (36%), Positives = 729/1192 (60%), Gaps = 25/1192 (2%)

Query: 1 MFLKEIELEGFKSFADKTKIEFDDKGVTA VVGPNPNSGKSNITESLRWALGESSAKNLRGGK 60

MFLK +++ GFKSFA++ ++F KGVTA VVGPNPNSGKSNIT+++RW LGE SA++LRGGK

Sbjct: 1 MFLKRLDVIGFKSFAERISVDFVKGTVTA VVGPNPNSGKSNITDAIRWVLGEQSARSIRGGK 60

Query: 61 MPDVI FAGTQNRNPLNYAKVAVVLDNSDHFIKTAKKEIRVERHIYRNGSDSYLIDGRKVR 120

M D+IFAG+ +R LN A+V + LDN DHF+ E+ V R +YR+G+S++LI+ + R

Sbjct: 61 MEDIIFAGSDSRKRLNLAEVTLTLDNDHFLPIDFHEVSVTRRVYRSGESEFLINNQPCR 120

Query: 121 LRDIHDLFMDTGLGRDSFSIISQGRVEEIFNSKPEERRAIFEEAAGVLKYKTRKKETQIK 180

L+DI DLFMD+GLG+++FSIISQG+VEEI +SK E+RR+IFEEAAGVLKYKTRKK+ + K

Sbjct: 121 LKDIIDLFDMSGLGKEAFSIIISQKVEEILSSKAEDRRSIFEEAAGVLKYKTRKKKAENK 180

Query: 181 LNQTQDNLDRLEDIIYELDTQLAPLEKQAKVAKQFLELDANRKQLQLDILVKDIDIAQER 240

L +TQDNL+R+EDI++EL+ Q+ PL+ QA +AK +LE + +++ + DI+ +

50 Sbjct: 181 LFETQDNLNRVEDILHELEGQVEPLKIQASTAKDYLEKKKELEHVEIALTAYDIEKLHGK 240

Query: 241 QTKDTEALAAALQQDLASYYAKRQSMEEYDQFKKQKQVLSQESDQTQTTLLELTKLIADL 300

+ E + +++ + + + E + + K Q L + ++ Q LL ++ + L

55 Sbjct: 241 WSTLKEKVQMAKEEELAESSAISAKEAKIEDTRDKIQALDES VNELQQVLLVTSEELEKL 300

Query: 301 EKQIELVKLESGQEAEEKAEAKKHLEQLQEQLDGFQAEKQCTEQQLH-----IDQQ 353

E + E++K E+K A ++ EQL+E + FQ +E E+L + ++

Sbjct: 301 EGRKEVLK-----ERKKNAVQNQEQLLEAIVQFQKQKQVLSQESDQTQTTLLELTKLIADL 353

Query: 354 CDVKQQLNELSNALERFSSDPDQLMETLREEFVLLMQEAAALSQLTALKAHLDKEKQAR 413

++ Q+ E AL + + ++ +E L+ ++ L+ +A++ N+L L + +

60 Sbjct: 354 KQLRAQVKEKQQLSLHNENVEEKIEQLKSDYFELLNSQASIRNELQLLDDQMSQSAVTL 413

Query: 414 QHKAQEQYQLLVTKLDQLNDESQKAQAHYKAQKEQVEMLLQNYQEGDKRVQELERDYQLNQ 473

Q A + + + ++ + + +++++ + Y++ + ++ +R Y+ N+

65

-453-

Sbjct: 414 QRLADNNEKHLQERHDISARKAACETEFARIEQEIHQSQVGAYRDMQTKYEQKKRQYEKNE 473  
 Query: 474 ERLFDLLDQKKGKEARKASLESIQKSHSQFYAGVRAVLQSQKLGIIIGAVSEHLSFSDSD 533  
 L+ + ++K LE++Q S FY GV+ VL++++LGGI GAV E +S +  
 5 Sbjct: 474 SALYQAYQYVQOARSKKDMLETMQGDFSGFYQGVKEVLKAKERLGGIRGAVLELISTEQK 533  
 Query: 534 YQTALEVALGANSQHIIVTDEAAAKRAIAYLKKNRQGRATFLPLTTIKARSLSEHYHRQL 593  
 Y+TA+E+ALGA++QH++ DE +A++AI YLK+N GRATFLPL+ I+ R L  
 10 Sbjct: 534 YETAIEIALGASAQHVVTDDDEQSARKAIQYLKQNSFGRATFLPLSVIRDRQLQSRDAETA 593  
 Query: 594 ATCEGYLGTAESLIRYDDSLSAIIONLLSSTAIFETIDQANIAARLLGYKVRIVTLDGTE 653  
 A +LG A L+ +D + ++IQNLL + I E + AN A+LLG++ RIVTL+G  
 Sbjct: 594 ARHSSFLGVASELVTFDPAYRSVIQNLGLTGLITEDLKGANELAKLLGHRYRIVTLEGDV 653  
 15 Query: 654 LRPGGSFSGGANRQSNNTTFI--KPELEQISEELTRLVEQLKITEKEVAALQSDLIKKEE 711  
 + PGGS +GGA ++ N + + ELE +++ L + E+ + E+EV L+ + +++  
 Sbjct: 654 VNPGGSMGTGAVKKKNNSSLGRSRELEDVTKRLAEMEETALLEQEVKTLKHSIQDMEKK 713  
 20 Query: 712 LTQLKLAGDQARLAEQ--RAQMAYQQLQEQEDSKALLAALDQSQTTHSDESLLAEQARI 769  
 L L+ G+ RL +Q + Q+ Q+ EK ++ L ++S + SDE + ++  
 Sbjct: 714 LADLRETGEGLRLKQDDVKQGLYELQVAEKNINTHELYDQEKSALESSEDEERKVRKRKL 773  
 Query: 770 EEALTAIAKKKNALTCDIDDIKENKDLIRQKTQNIHQALSQARLQERDLLNEKKFEQANQ 829  
 EE L+A+++K L DID + + K +++ L++ ++ K E+ N  
 25 Sbjct: 774 EEELSAVSEKMKQLEEDIDRLTKQKQTSSTKESLSNELTELKIAAAKKEQACKGEEDNL 833  
 Query: 830 SRLRTQLKQCQONILKLESILNNNVSDSIQRLPQWQKQLQDATEHKSGAQKRLVQLRFE 889  
 +RL+ +L + + + + + L+ S+ S +++L++A +HK + + ++L  
 30 Sbjct: 834 ARLKKELTETELALKEAKEDLSFLTSEMSSS--TSGEEKLEEAAKHKLNDKTKTIELIAL 891  
 Query: 890 IEDYEARLEETAEKITKESEKNDTFIRRQTKL---ETHLEQVANRLRAYAKSLSEDFQM 945  
 D +L+ + +E ++ +++T L E L ++ L + L E++ +  
 Sbjct: 892 RRDQRIKLQHGLDITYERELKEMKRLYKQKTLLKDEEVKLGRMEVELDNLQYLREEYSL 951  
 35 Query: 946 TLADAKEVNTNSIDHLESACEKLHHLQKTIRALGPINSDAINQYEEVHERLTFLTTSQKTDL 1005  
 + AKE E A++++ ++ I LG +N +I+++E V+ER FL+ QK DL  
 Sbjct: 952 SFEGAKEKYQLETDPEEARKRVKLIKLAIEELGTVNLGSIDEFERVNERVKFLSEQKEDL 1011  
 40 Query: 1006 TKAKNLLLETINSMDSEVKARFKVTFEAIQKSFKETFTQMFGGGSADLVLTE-TDILLSAG 1064  
 T+AKN L + I MD E+ RF TF I+ F + F +FGGG A+L LT+ DLL +G  
 Sbjct: 1012 TEAKNTLFPQVIEEMDEEMTKRFNDTFVQIRSHFDQVFRSLFGGGAELRLTDPNDLLHSG 1071  
 Query: 1065 IEISVQPPGKKIQSLNLMMSGGEKALSALALLFAIIRVKTIPFVILDEVEAALDEANVKRF 1124  
 +EI QPPGKK+Q+LNL+SGGE+AL+A+ALLF+I++V+ +PF +LDEVEAALDEANV RF  
 45 Sbjct: 1072 VEIIAQPPGKKLQNLNLLSGGERALTAIALLFSILKVRPVPFCVLDEVEAALDEANVRF 1131  
 Query: 1125 GDFLNRFDKDSQFIVVTHRKGTMMAADSIYGITMQESGVSKIVSVKLKEAQE 1176  
 +L ++ D+QFIV+THRKGTM AD +YG+TMQESGVSK++SVKL+E +E  
 50 Sbjct: 1132 AQYLKYYSSDTQFIVITHRKGTMEEADVLYGVTMQESGVSKIVSVKLEETKE 1183

An alignment of the GAS and GBS proteins is shown below:

Identities = 732/1179 (62%), Positives = 911/1179 (77%)

55 Query: 1 MFLKEIEMQGFKSFADKTKVEFDQGVTAUVGPNNGSGKSNITESLRWALGESSAKSLRGGK 60  
 MFLKEIE++GFKSFADKTK+EFD+GVTAVVGPNGSGKSNITESLRWALGESSAK+LRGGK  
 Sbjct: 1 MFLKEIELEGFKSFADKTKIEFDKGVTAUVGPNNGSGKSNITESLRWALGESSAKNLRGGK 60  
 Query: 61 MPDVIFAGTENRKPLNYAQVSVTLDNSDHFIENIADEVRRVERIFRNGDSEYLDGRKVR 120  
 MPDVIFAGT+NR PLNYA+V+V LDNSDHFI+ E+RVER I+RNGDS+YLDGRKVR  
 60 Sbjct: 61 MPDVIFAGTQNRNPLNYAKVAVVLDNSDHFIKTAKKEIRVERHIYRNGDSYLDGRKVR 120  
 Query: 121 LRDIHDLFMDTGLGRDSFSIISQGRVEAIFNSKPEERRAIFEEAAGVLKYKTRKKETQSK 180  
 LRDIHDLFMDTGLGRDSFSIISQGRVE IFNSKPEERRAIFEEAAGVLKYKTRKKETQ K  
 Sbjct: 121 LRDIHDLFMDTGLGRDSFSIISQGRVEEIFNSKPEERRAIFEEAAGVLKYKTRKKETQIK 180  
 65 Query: 181 LEQTQGNLDRLIEDIYELDMQVPLEKQASIAKRFVLVDEERQGLHLSILIEDILQHQS 240  
 L QTQ NLDRLIEDIYELD Q+ PLEKQA +AK+FL LD R+ L L IL++DI Q

Sbjct: 181 LNQTQDNLDRLIEDIYELDTQLAPLEKQAKVAKQFLELDANRKQLQLDILVKDIDIAQER 240  
 Query: 241 LTTVEEKLTVRKELATYYQQRQSLDENQSLKQKRHLSEEIEAKQLALLDVTKLKSDL 300  
 T E L ++++LA+YY +RQS+E++ Q KQK+ LS+E + Q LL++TKL +DL  
 5 Sbjct: 241 QTKDTEALALQQDLASYAKRQSMEDYQKFKQKQVLSQESDQTQTTLLELTCLIADL 300  
 Query: 301 ERQIDLIRLESNQAEKKEEAGQRLAELEIKAKDCSDQITQKNIELTTLSEKIAQIRSEI 360  
 E+QI+L++LES Q+AEKK EA + L +L+ + + Q +L + +++ ++ ++  
 10 Sbjct: 301 EKQIELVKLESQGAEEKKAEAKHLEQLQEQLDGFQAEKQCTEQLLHIDQQLCDVKQQL 360  
 Query: 361 VSTESSLERFSTNPDQIIEKLREDFVTLMQEEADTSNALTALLADIENQKQASQAKSQEI 420  
 ++LERFS++PDQ++E LRE+FV LMQ+EA SN LTAL A ++ +KQA Q K+QE  
 Sbjct: 361 NELSNALERFSSDPDQMETLREEFVLLMQKEAALSNQLTALKAHLDKEKQARQHKAQEY 420  
 15 Query: 421 QEVSKNLEVLKSNAKVALERFEAAKNVRQLLSHYQDLGQTLQNLGEYKNQQSILFDHL 480  
 Q + L+ L ++ A ++A K+ V LL +YQ+ + +Q LE +Y+ Q LFD L  
 Sbjct: 421 QLLVTKLDQLNDESQKAQAHYKAQKEQVEMLLQNYQEGDKRVQELERDYQLNQERLFDLL 480  
 20 Query: 481 DEIKSKQARISSLESILKNHSNFYAGVKSVLQAKDQLGGIIGAVSEHLSFDKHYQTALEI 540  
 D+ K K+AR +SLESI K+HS FYAGV++VLQ++ +LGGIIGAVSEHLSFD YQTALE+  
 Sbjct: 481 DQKKGKEARKASLESIQKSHSQFYAGVRAVLQSQKKLGGIIGAVSEHLSFSDYQTALEV 540  
 Query: 541 ALGSSQHIIVEDESAAKRSIAFLKKNRQGRATFLPLTTIKPRELAQHLYSLQSSQGFL 600  
 ALG +SQHIIV DE+AAKR+IA+LKKNRQGRATFLPLTTIK R L++HY +L + +G+L  
 25 Sbjct: 541 ALGANSQHIIVTDEAAAKRAIAYLKKNRQGRATFLPLTTIKARSLSEHYHRQLATCEGYL 600  
 Query: 601 GIASELVTYDQRLSNIFKNLGLTAIFDVTDNANVAARQLNYQVRLVTLTGTELTPGGSY 660  
 G A L+ YD LS I +N L TAIF+T+D AN+AAR L Y+VR+VTLTGTELTPGGS+  
 30 Sbjct: 601 GTABSLIRYDDSLSAIIQNLLSTAIFETIDQANIAARLLGYKVRIVTLTGTELTPGGSF 660  
 Query: 661 SGGANRQNNTVFIKPELDNLKKELKQAQSKQLIQEKEVATLLEQLKEKQETLAQLKNDGE 720  
 SGGANRQ+NT FIKPEL+ + +EL + + I EKEVA L L K+E L QLK G+  
 Sbjct: 661 SGGANRQSNNTTFIKPELEQISEELTRLVEQLKITEKEVAALQSDLIKKEELTQLKLAGD 720  
 35 Query: 721 QARLEQRADIEYQQLSEKLADLNKLYNGLQLSSGALEQTTSENEKNRLEKELEQFAIKK 780  
 QARL EQRA + YQQL EK D L L S + E+ R+E+ L A KK  
 Sbjct: 721 QARLAEQRAQMAYQQLQEQEDSKALLAALDQSQTTTHSDESLAEQARIEEALTAIAKKK 780  
 40 Query: 781 EELTTSIAQIKEDKDSIQEKVNNTLTLSEAQLERDLLENEQKFERANCTRLEITLSEIK 840  
 LT I IKE+KD I++K N+ LS+A+L+ERDLLENE+KFE+AN +RL L + +  
 Sbjct: 781 NALTCDIDDIKENKDLIRQKTQNIHQALSQARLQERDLLENEKKFEQANQSRLTQLKQCQ 840  
 Query: 841 RDISNLQTLTSHQDSQLDKELPRIEKQLLQVNNRRENDEEKLVSRLFELEDCEAALDDL 900  
 ++I L+++L++ SQ + LP+ +KQL + +++LV LRFE+ED EA L++  
 45 Sbjct: 841 QNILKLESILNNVVSQDSIQRLPQWQKQLQDATEHKSGAQKRLVQLRFEIEDYEARELEET 900  
 Query: 901 AASLAKEGQKNESLIRQQAQLESQCEQLSQQLMIFSRQLSQEDYQMTLDEAKVKANVLEDI 960  
 A + KE +KN++ IR+Q +LE+ EQ++ +L +++ LSED+QMTL +AK N ++ +  
 50 Sbjct: 901 AEKITKESEKNDTFIRROTLETHLEQVANRLRAYAKSLSEDFQMTLDAKEVTNSIDHL 960  
 Query: 961 LMAREQLKSLQAKIKALGPVNIDAIAQFEEVHERLTFLNTQRDDLVAKNLLETTITDMD 1020  
 A+E+L LQ I+ALGP+N DAI Q+EEVHERLTFL +Q+ DL AKNLLETTI MD  
 Sbjct: 961 ESAKEKLHHLQKTIRALGPINSDAINQYEEVHERLTFLTSQKTDLTAKKNLLETTINSMD 1020  
 55 Query: 1021 DEVKTRFKSTFEAIRHSFKETFVQMFGGGSADLILTEGDLLSAGVDISVQPPGKKIQSLN 1080  
 EVK RPK TFEAI+ SFKETF QMFGGGSADL+LTE DLLSAG++ISVQPPGKKIQSLN  
 Sbjct: 1021 SEVKARFKVTFEAIQKSFKETFTQMFGGGSADLVLTETDILLSAGIEISVQPPGKKIQSLN 1080  
 60 Query: 1081 LMSGGEKALSALALLFAIRVKTIPIFVILDEVEAALDEANVKRFGDYLNRFDKSSQFIVV 1140  
 LMSGGEKALSALALLFAIRVKTIPIFVILDEVEAALDEANVKRFGD+LNRFDK SQFIVV  
 Sbjct: 1081 LMSGGEKALSALALLFAIRVKTIPIFVILDEVEAALDEANVKRFGDFLNRFDKSSQFIVV 1140  
 Query: 1141 THRKGTMASADSIYGVTMQESGVSKIVSVKLKEAQEMTN 1179  
 THRKGTM+AADSIYG+TMQESGVSKIVSVKLKEAQEMTN  
 65 Sbjct: 1141 THRKGTMASADSIYGITMQESGVSKIVSVKLKEAQEMTN 1179



-455-

SEQ ID 1162 (GBS199) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 2; MW 75kDa).

GBS199-GST was purified as shown in Figure 208, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 359

A DNA sequence (GBSx0390) was identified in *S.agalactiae* <SEQ ID 1165> which encodes the amino acid sequence <SEQ ID 1166>. This protein is predicted to be ribonuclease III (*rnc*). Analysis of this protein sequence reveals the following:

```

Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3372(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9711> which encodes amino acid sequence <SEQ ID 9712> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CAB13466 GB:Z99112 ribonuclease III [Bacillus subtilis]
Identities = 115/230 (50%), Positives = 154/230 (66%), Gaps = 1/230 (0%)

Query: 13 KKMKELRSKLEKDYGIVFANQELLDTAFTHTSYANEHRLNLISHNERLEFLGDAVLQLLI 72
KK+++ + E+ + F N++LL AFTH+SY NEHR NERLEFLGDAVL+L I
Sbjct: 15 KKVEQKFKEFQER-ISVHFQNEKLLYQAFTHSSSYVNEHRKKPYEDNERLEFLGDAVLELTI 73

Query: 73 SQYLF TKYPQKABGDL SKLRSMIVREESLAGFSRLCGFDHYIKLGKGEESGGRNRDTIL 132
S++LF KYP +EGDL+KLR+ IV E SL + F + LGKGEE +GGR R +L
Sbjct: 74 SRFLFAKYPAMSEGD LTKLRAAIVCEPSLVSLAHEL SFGDLVLLGKGEEMTGGRKRPALL 133

Query: 133 GDLFEAF LGALLLDKGV E VVHAFVNKMIPHVEKGT YERVKDYKTS LQELLQSHGDVKID 192
D+FEAF+GAL LD+G+E V +F+ + P + G + V D+K+ LQE +Q G ++
Sbjct: 134 ADVFEAF IGALYLDQGLEPVSFSLKVIVFPKINDGAFSHVMDFKSQLQEYVQRD GKGSL E 193

Query: 193 YQVTNESGPAHAKEFEVTVSVNQENLSQGIGRSKKA AEQDAAKNALATLQ 242
Y+++NE GPAH +EFE VS+ E L G GRSKK AEQ AA+ ALA LQ
Sbjct: 194 YKISNEKGPAHNREFEAI VSLKGEPLGVGNRSKKEAEQHAAQEALAKLQ 243

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1167> which encodes the amino acid sequence <SEQ ID 1168>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1414(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 170/227 (74%), Positives = 192/227 (83%)

Query: 15 MKELRSKLEKDYGIVFANQELLDTAFTHTSYANEHRLNLISHNERLEFLGDAVLQLLISQ 74

```

-456-

MK+L L + I F + LL+TAFTHTSYANEHRLN+SHNERLEFLGDAVLQL+IS+  
 Sbjct: 1 MKQLEELLSTSFDIQFNDLTLLETAFTHTSYANEHRLNVSHNERLEFLGDAVLQLIISE 60  
 Query: 75 YLFTKYPQKAEGDLSKLRSIMIVREESLAGFSRLCGFDHYIKLGKGEKSGGRNRDTILGD 134  
 YLF KYP+K EGD+SKLRSIMIVREESLAGFSR C FD YIKLGKGEKSGGR RDTILGD  
 Sbjct: 61 YLFAKYPKKTGDMSKLRSIMIVREESLAGFSRFSFDAYIKLGKGEKSGGRRRDTILGD 120  
 Query: 135 LFEAFILGALLLDKGVEVVHAFVNKVMIPHVEKGTIERVKDYKTSLQELLQSHGDVKIDYQ 194  
 LFEAFILGALLLDKG++ V F+ +VMIP VEKG +ERVKDYKT LQE LQ+ GDV.IDYQ  
 Sbjct: 121 LFEAFILGALLLDKGIDAVRRFLKQVMIPQVEKGNFERVKDYKTCLQEFLQTKGDVAIDYQ 180  
 Query: 195 VTNESGPAHAKEFEVTVSVNQENLSQIGRSKKAEEQDAAKNALATL 241  
 V +E GPAHAK+FEV++ VN LS+G+G+SKK AEQDAAKNALA L  
 Sbjct: 181 VISEKGAHAKEFEVSIVVNGAVLSKGLGKSKLAEEQDAAKNALAQL 227

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 360

A DNA sequence (GBSx0391) was identified in *S.agalactiae* <SEQ ID 1169> which encodes the amino acid sequence <SEQ ID 1170>. Analysis of this protein sequence reveals the following:

Possible site: 43  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -4.19 Transmembrane 100 - 116 ( 99 - 117)  
 INTEGRAL Likelihood = -2.44 Transmembrane 81 - 97 ( 81 - 97)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2678(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAC12789 GB:AJ279090 hypothetical protein [Staphylococcus  
 carnosus]  
 Identities = 50/114 (43%), Positives = 72/114 (62%)  
 Query: 3 KIFYISLGFISLGIGIAGIVLPVPTTLPVLLSAFCFSRSEKFDIWLRTKVYKYAAD 62  
 K ++LG I GIG GIV+P++PTTP +LL+A CFSRSS+KF+ WL TK++ Y  
 Sbjct: 2 KYVLMTLGLIFAGIGFVGIVVPLPTTPFLLLAICFSRSSKKFNRWLVTNKIHDEYVES 61  
 Query: 63 FVESRSIAPARKKSMIWIYILMGISIIYFAPLMWLKGLLIGTIVGTIVLFYV 116  
 F + +K ++ +YILMGISI+ +++++ LLI V T VLF V  
 Sbjct: 62 FKRDKGFTLKKFKLLTSLYILMGISIFIIDNLYIRITLLIMLFVQTVVLFTEFV 115

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 361

A DNA sequence (GBSx0392) was identified in *S.agalactiae* <SEQ ID 1171> which encodes the amino acid sequence <SEQ ID 1172>. Analysis of this protein sequence reveals the following:

Possible site: 45  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1908(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-457-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1173> which encodes the amino acid sequence <SEQ ID 1174>. Analysis of this protein sequence reveals the following:

Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1610(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 225/269 (83%), Positives = 248/269 (91%)

Query: 1 MSEIGFKYSILASGSTGNCIFYETPQKRLLIDAGLTGKKVTSLLAEINRKPEDLDAILVT 60  
M+E GFKYSILASGSTGNCFY+ETP+KRLIDAGLTGKK+TSLLAEI+RKPEDLDAIL+T  
Sbjct: 1 MNESGFKYSILASGSTGNCFYLET PKRLLIDAGLTGKKITSLLAEIDRKPEDLDAILIT 60

Query: 61 HEHSDHIKGVGV LARKYHLDIYANEQTWKVMDERNMLGKVDVSQKHVFGRGKTLTFGDLD 120  
HEHSDHIKGVGV+ARKYHLDIYANE+TW++MDE NMLGK+D SQKH+F R K LTFGD+D  
Sbjct: 61 HEHSDHIKGVGV MARKYHLDIYANEKTWQLMDECNMLGKLDASQKHIFQRDKVLTFGDVD 120

Query: 121 IESFGVSHDAVDPQFYRMKDDKSFVMLTDTGYVSDRMAGLIENADGYLIESNHDIEILR 180  
IESFGVSHDA+DPQFYR+MKD+KSFVMLTDTGYVSDRM G+IENADGYLIESNHDIEILR  
Sbjct: 121 IESFGVSHDAIDPQFYRIMKDNKSFVMLTDTGYVSDRMTGIENADGYLIESNHDIEILR 180

Query: 181 SGSYPWTLKQRI LSKGHL SNEDGSETMIRTIGNRTKHIYLGHL SKENNIKELAHMTMEN 240  
SGSYPW+LKQRI LSKGHL SNEDG+ MIR++G TK IYLGHL SKENNIKELAHMTM N  
Sbjct: 181 SGSYPWSLQRI LSKGHL SNEDGAGAMIRSLGYNTKKIYLGHL SKENNIKELAHMTMVN 240

Query: 241 NLMRADFGVGTDFSVHDTSPDSATPLTRI 269  
L AD VGTDF+VHDTSPD+A PLT I  
Sbjct: 241 QLAMADLAVGTDFTVHDTSPDTACPLTDI 269

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 362

A DNA sequence (GBSx0393) was identified in *S.agalactiae* <SEQ ID 1175> which encodes the amino acid sequence <SEQ ID 1176>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -11.94 Transmembrane 15 - 31 ( 5 - 34)

----- Final Results -----

bacterial membrane --- Certainty=0.5776(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1177> which encodes the amino acid sequence <SEQ ID 1178>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-458-

An alignment of the GAS and GBS proteins is shown below:

Identities = 335/443 (75%), Positives = 392/443 (87%)

```

5   Query: 7   NIRSFELALLFLLVFVAVYFVYLAVRDFKMSKNIRLLNWKVRDLIAGNYSILIQGDAD 66
      N+ +FELA+L LLVFVA YF++LAVRD++ ++ IR+++ K+RDLI G Y+D I + D +
      Sbjct: 8   NLSTFELAILILLVFVAFYFIHLAVRDYRNARIIRMMSHKIRDLINGRYTDIIDEKADIE 67

10  Query: 67   LVELGESLNDLSDVFRMAHDNLEQEKNLASILTYMTDGVLATDRSGKIVMINETAQQQF 126
      L+EL + LNDLSDVFR+ H+NL QEKNRLASIL YM+DGVLATDRSGKI+MINETA++Q
      Sbjct: 68   LMELSDQLNDLSDVFRILTHENLAQEKNLASILAYMSDGVLATDRSGKIIMINETARKQL 127

      Query: 127  NLAYDEALSMNIVDMLGSGSPYSFQDLVSKTPEVVLNRRDENGFEVTLRIRFALNRRESG 186
      NL+ +EAL NI D+L + Y+++DLVSKTP V +N R++ GEFV+LR+RFALNRRESG
15  Sbjct: 128  NLSKEEALKKNITDLEGGDTSYTYRDLVSKTPVVTVNSRNDMGEFVSLRLRFALNRRESG 187

      Query: 187  FISGLVAVSHDATEQEKEERERRLFVSNVSHELRTPLTSVKSYLEALDEGALNEEVAPSF 246
      FISGLV V HD TEQEKEERERRLFVSNVSHELRTPLTSVKSYLEALDEGAL E++APSF
      Sbjct: 188  FISGLVVVLDHDTTEQEKEERERRLFVSNVSHELRTPLTSVKSYLEALDEGALKEDIAPSF 247

20  Query: 247  IKVSLDETNRMMRMISDLLSLSRIDNEVTHLDVEMTNFTAFMTSILNRFDQIRNQKTVTG 306
      IKVSLDETNRMMRMISDLL+LSRIDN+VT L VEMTNFTAF+TSILNRFD ++NQ T TG
      Sbjct: 248  IKVSLDETNRMMRMISDLLNLSRIDNQVTQLAVEMTNFTAFITSILNRFDLVKNQHTGTG 307

25  Query: 307  KVEIVRDYPLKSIWVEIDTDKMTQVIDNILNNAVKYSPDGGKITVNLRTTKTQMILSIS 366
      KVEIVRDYP+ S+W+EID DKMTQVI+NILNNA+KYSPDGGKITV ++TT TQ+I+SIS
      Sbjct: 308  KVEIVRDYPITSVWIEIDNDKMTQVIENILNNAIKYSPDGGKITVRMKTDTQLIISIS 367

      Query: 367  DQGLGIPKDKDLPLIFDRFYRVDKARSRKQGGTGLGLSIAKEIVKQHKGFIWAKSEYKGKS 426
      DQGLGIPK DLPLIFDRFYRVDKARSR QGGTGLGL+IAKEI+KQH GFIWAKS+YGKGS
30  Sbjct: 368  DQGLGIPKTDLPLIFDRFYRVDKARSRAQGGTGLGLAIAKEIKQHKGFIWAKSDYKGKS 427

      Query: 427  TFTIVLPYDKDAVTYEEWEDVED 449
      TFTIVLPY+KDA YEEWE+ D
35  Sbjct: 428  TFTIVLPYEKDAAIYEEWEEDVD 450

```

A related GBS gene <SEQ ID 8561> and protein <SEQ ID 8562> were also identified. Analysis of this protein sequence reveals the following:

```

40  Lipop: Possible site: -1   Crend: 8
      McG: Discrim Score:      8.59
      GvH: Signal Score (-7.5): -3.38
          Possible site: 26
      >>> Seems to have an uncleavable N-term signal seq
      ALOM program   count: 1 value: -11.94 threshold: 0.0
45  INTEGRAL    Likelihood = -11.94   Transmembrane   15 - 31 ( 5 - 34)
      PERIPHERAL Likelihood = 8.27     178
      modified ALOM score: 2.89

      *** Reasoning Step: 3

50  ----- Final Results -----
          bacterial membrane --- Certainty=0.5776(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
55

```

The protein has homology with the following sequences in the databases:

```

60  67.5/83.5% over 439aa
      Streptococcus pneumoniae
      GP|5830524| histidine kinase Insert characterized

      ORF01458(331 - 1647 of 1947)
      GP|5830524|emb|CAB54569.1|AJ006392(10 - 449 of 449) histidine kinase {Streptococcus
      pneumoniae}
      %Match = 45.6

```

-459-

%Identity = 67.5 %Similarity = 83.4

Matches = 297 Mismatches = 70 Conservative Sub.s = 70

```

5      126      156      186      216      246      276      306      336
      ITSPFSDTYRTSHDTRTFIGNSLGI*LFWRCPYS*CDGETFT*KD*RYSWSSRIYFDSTWCRXIT*SLMNNSAANIRSF
                                         |
                                         MLDLLKQTIFT
                                         10

10     366      396      426      450      480      510      540      570
      LALLFLLVFVAVYFVYLAVRDFKMSKNIRL--LNWKVRDLIAGNYSILIQGDADLVELGESINDLSDFVRMAHDNLEQ
      ::||:::  :|  :  ||:|  :|  ||:||||:|  :||  ::  :  :|||:|  |::|||
      RDPFIFILILLGFILVVTLLENNRRDNIQLKQVNQKVKDLIAGDYSKVLDMQGSGSEITNITNNLNDLSEVIRLTQENLEQ
                                         30      40      50      60      70      80      90

15     600      630      660      690      720      750      780      810
      EKNRLASILTYMTDGVLATDRSGKIVMINETAQQQFNLAYDEALSMNIVDMLGSGSPYSFQDLVSKTPEVVLNRRDENG
      |  ||  |||  |||:|:|  :|:|:|:|:|  :|  :|  :|:|  :|  :|:|:|:|:|:|  :|  |||
      ESKRLNSILFYMTDGVLATNRRGQIIMINDTAKQLGLVKEDVLNRSILELLKIEENVELRDLITQSPPELLDSQDINGE
                                         110     120     130     140     150     160     170

20     840      870      900      930      960      990      1020     1050
      FVTLRIRFALNRRESGFISGLVAVSHDATEQEKEERERRLFVSNVSHELRTPLTSVKSYLEALDEGALNEEVAPSFIVK
      ::  ||:|  |||:|:|  |||:|:|  |||:|:|  |||:|:|  |||:|:|  |||:|:|  |||:|:|  |||:|:|
25     YLNLRVRFALIRRESGFISGLVAVLHDTTEQEKEERERRLFVSNVSHELRTPLTSVKSYLEALDEGALCETVAPDFIVK
                                         190     200     210     220     230     240     250

30     1080     1110     1140     1170     1200     1230     1260     1290
      LDETNRMMRMISDLLSLSRIDNEVTHLDVEMTNFTAFMTSILNRFQIRNQKTVTGKVEIVRDYPLKSIWVEIDTDKMT
      |||:|:|:|:|:|  |||:|:|  :|||:|  |||:|  |||:|:|  :|  |||:|:|  :|  |||:|:|  :|  |||:|:|
      LDETNRMMRMVTDLLHLSRIDNATSHLDVELINFTAFITFILNRFDMKGQ--EKEKYELVRDYPINSIWMEIDTDKMT
                                         270     280     290     300     310     320

35     1320     1350     1380     1410     1440     1470     1500     1530
      QVIDNILNNAVKYSPDGGKITVNLRTTKQMILSISDQGLGIPKDLPLIFDRFYRVDKARSRKQGGTGLGLSIAKEIVK
      ||:|:|:|:|:|  |||:|:|  :||:|  |||:|:|:|:|:|  |||:|:|  |||:|:|  |||:|:|  |||:|:|
      QVVDNILNNAIKYSPDGGKITVRMKTTEQMIILSISDHGLGIPKQDLPRIFDRFYRVDKARSRAQGGTGLGLSIAKEIIK
                                         340     350     360     370     380     390     400

40     1560     1590     1620     1647     1677     1707     1737     1767
      QHKGFIWAKSEYKGSTFTIVLPYDKDAVTYEWEW-VED*NMSEIGFKYSILASGSGTNCIFYIETPQKRLIDAGLTGK
      |||:|:|:|:|:|  |||:|:|  |||  |||
      QHKGFIWAKSEYKGSTFTIVLPYDKDAVKEEVEWEVED
                                         420     430     440

```

SEQ ID 1176 (GBS41) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 7; MW 50kDa), in Figure 168 (lane 2-4; MW 65kDa – thioredoxin fusion) and in Figure 238 (lane 4; MW 65kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 7; MW 75kDa).

Purified Thio-GBS41-His is shown in Figure 244, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 363

A DNA sequence (GBSx0394) was identified in *S. agalactiae* <SEQ ID 1179> which encodes the amino acid sequence <SEQ ID 1180>. This protein is predicted to be VicR protein (regX3). Analysis of this protein sequence reveals the following:

Possible site: 60

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

-460-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2754(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1181> which encodes the amino acid sequence <SEQ ID 1182>. Analysis of this protein sequence reveals the following:

Possible site: 60

10 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.2754(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 205/236 (86%), Positives = 221/236 (92%)

20 Query: 1 MKKILIVDDEKPISDIIKFNLTKEGYETATAFDGREALVQYAEFQPDLIILDMLPELDG 60  
 MKKILIVDDEKPISDIIKFNLTKEGY+ TAFDGRE+ + E +PDLIILDMLPELDG  
 Sbjct: 1 MKKILIVDDEKPISDIIKFNLTKEGYDIVTAFDGREAVTIFEEKPDLIILDMLPELDG 60

25 Query: 61 LEVAKEVRKTSHPIIMLSAKDSEFDKVIGLEIGADDYVTKPFSNRELLARVKAHLRRT 120  
 LEVAKE+RKTSH+PIIMLSAKDSEFDKVIGLEIGADDYVTKPFSNRELLARVKAHLRRT  
 Sbjct: 61 LEVAKEIRKTSHPVPIIMLSAKDSEFDKVIGLEIGADDYVTKPFSNRELLARVKAHLRRT 120

30 Query: 121 NIETAVAEESAQNASSDITIGELQILPDFAFIKKRGEEIELTHREFELLHHLATHIGQVM 180  
 IETAVAE+ + + +TIG LQILPDAF+AKK G+E+ELTHREFELLHHLA H+GQVM  
 Sbjct: 121 TIETAVAEENASSGTQELTIGNLQILPDAFVAKKHGQEVETLTHREFELLHHLANHMGQVM 180

35 Query: 181 TREHLLLETVWGYDYFGDVRTVDVTVRRLREKIEDTPGRPEYILTRRGVGYMKSYE 236  
 TREHLLLE VWGYDYFGDVRTVDVTVRRLREKIEDTP RPEYILTRRGVGYMKSY+  
 Sbjct: 181 TREHLLLEIVWGYDYFGDVRTVDVTVRRLREKIEDTPSRPEYILTRRGVGYMKSYD 236

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 364

40 A DNA sequence (GBSx0395) was identified in *S.agalactiae* <SEQ ID 1183> which encodes the amino acid sequence <SEQ ID 1184>. This protein is predicted to be amino acid ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3791(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14701 GB:Z99118 glutamine ABC transporter (ATP-binding protein) [Bacillus subtilis]

Identities = 149/244 (61%), Positives = 200/244 (81%), Gaps = 2/244 (0%)

55 Query: 3 LISYKVNKYYG DYHALRQINLEIEPGQVVVLLGPSGSGKSTLIRTMNALESIDGSLVV 62  
 +I+++NVNK+YGD+H L+QINL+IE G+VVV++GPSGSGKSTL+R +N LESI++G L V  
 Sbjct: 1 MITFQNVNKH YGDFHVLKQINLQIEKGEVVVVIIGPSGSGKSTLLRCINRLESINEGVLT V 60

-461-

Query: 63 NGHELANISSKELVNLKKEVGMVFQHFNLPHKTVLENITLAPIKVLKQSKKEAMEIAEK 122  
 NG + N ++ +R+ +GMVFQHF+LYPHKTVL+NI LAP+KVL+QS ++A E A  
 Sbjet: 61 NGTAI-NDRKTDINQVRQNIQGMVFQHFHLYPHKTVLQNI MLAPVKVLRQSPQAKETARY 119

5 Query: 123 YLKFMNMWERKDSYPSMLSGGQKQRIAIARGLAMHPKLLLFDEPTSALDPETIGDVL SVM 182  
 YL+ V + ++ D+YPS LSGGQ+QR+AIARGLAM P+++LFDEPTSALDPE IG+VL VM  
 Sbjet: 120 YLEKVGIPDKADAYPSQLSGGQQQVAIARGLAMKPEVMLFDEPTSALDPEMIGEVL DVM 179

10 Query: 183 QKLANDGMNMVVVTHEMGFAREVADRIIFMADGEILVDTTDVQDFFDNPREPRAKQFLSN 242  
 + LA +GM MVVVTHEMGFA+EVADRI+F+ +G+IL + +F+ NP+E RA+ FLS  
 Sbjet: 180 KTLAKEGMTMVVVTHEMGFAKEVADRIVFIDEKGILEEAVPA-EFYANPKERARLFLSR 238

Query: 243 IINH 246  
 I+NH  
 15 Sbjet: 239 ILNH 242

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1185> which encodes the amino acid sequence <SEQ ID 1186>. Analysis of this protein sequence reveals the following:

Possible site: 51  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3763(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 25 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 131/243 (53%), Positives = 179/243 (72%), Gaps = 2/243 (0%)

30 Query: 2 SLISYKNVKNYYGDYHALRQINLEIEPGQVVVLLGPSGSGKSTLIRTMNALESIDDGSLV 61  
 ++IS K+++KYYG L+ I+L+I PG+VVV++GPSGSGKSTL+RTMN LE G +  
 Sbjet: 5 AIIISKDLHKYGHNEVLKGIDLDIMPGEVVVIIGPSGSGKSTLLRTMNLLEVPTKGQIR 64

35 Query: 62 VNGHELANISSKELVNLKKEVGMVFQHFNLPHKTVLENITLAPIKVLKQSKKEAMEIAE 121  
 G ++ + ++ ++R+++GMVFQ FNL+P+ T+LENITL+PIK +K EA + A  
 Sbjet: 65 FEGIDITD-KKNDIFSMREKMGGMVFQGFNLFPNMILENITLSPIKTKGMAKAEADKTAL 123

40 Query: 122 KYLKFMNMWERKDSYPSMLSGGQKQRIAIARGLAMHPKLLLFDEPTSALDPETIGDVL SVM 181  
 L V + E+ +YP+ LSGGQ+QRIAIARGLAM P +LLFDEPTSALDPE +G+VL+V  
 Sbjet: 124 SLLDKVGLSEKAKAYPASLSGGQQQRIAIARGLAMPDVLVLFDEPTSALDPEMVGEVLAV 183

45 Query: 182 MQKLANDGMNMVVVTHEMGFAREVADRIIFMADGEILVDTTDVQDFFDNPREPRAKQFLS 241  
 MQ LA GM MV+VTHEMGFA+EVADR++FM DG ++V+ FD +E R K FLS  
 Sbjet: 184 MQDLAKSGMTMVIVTHEMGFAKEVADRVFM-DGGVIVEEGSPNQLFDLTKEERTKDFLS 242

Query: 242 NII 244  
 ++  
 Sbjet: 243 RVL 245

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 365

A DNA sequence (GBSx0396) was identified in *S.agalactiae* <SEQ ID 1187> which encodes the amino acid sequence <SEQ ID 1188>. This protein is predicted to be glutamine-binding. Analysis of this protein sequence reveals the following:

Possible site: 27  
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

-462-

bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB73178 GB:AL139076 probable ABC-type amino-acid transporter  
 periplasmic solute-binding protein [Campylobacter  
 jejuni]

Identities = 99/240 (41%), Positives = 141/240 (58%), Gaps = 3/240 (1%)

10

Query: 1 MLRRKRLTFYLLSCIFILFLYPNSTSANQLSEIKSGVLKVGKQDVPNFGYYNAETNQ 60  
 M+ RK L + + + F + + +L IK G L VGK DVP++ + T +  
 Sbjct: 1 MVFRKSLKLAVFALGACVAFSNANAEGKLESIKSGQLIVGVKNDVPHYALLDQATGE 60

15

Query: 61 YEGMEIDIAKKIAKSL---GVKPVFVPTTAQTREPLMDNGQIDILIATYTTITPERKANYN 117  
 +G E+D+AK +AKS+ K V A+TR PL+DNG +D +IAT+TITPERK YN  
 Sbjct: 61 IKGFEVDVAKLLAKSILGDDKKIKLVAVNAKTRGPLLDNGSVDAVIATFTTITPERKRIYN 120

20

Query: 118 ISKAYYHDEIGFLVRKNSHIKTIKELDGKHIGVAQGATTKVNLEKYAKEHKLKFSYAQLG 177  
 S+ YY D IG LV K K++ ++ G +IGVAQ ATTK + + AK+ + +++  
 Sbjct: 121 FSEPPYQDAIGLLVLKEKKYKSLADMKGANIGVAQAATTKKAIGEAAKKIGIDVKFSEFP 180

25

Query: 178 SFPELAISLYANRIDAFSVDKSILSGYLSPHTTILKEGFNTQYGIATSKQDKVLIPYVN 237  
 +P + +L A R+DAFSVDKSIL GY+ + IL + F Q YGI T K D YV+  
 Sbjct: 181 DYPsIKAALDAKRVDAFSVDKSILLGVDDKSEILPDSFEPQSYGIVTKDDPAFAKYVD 240

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1189> which encodes the amino acid sequence <SEQ ID 1190>. Analysis of this protein sequence reveals the following:

30

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -6.16 Transmembrane 17 - 33 ( 15 - 35)

35

----- Final Results -----  
 bacterial membrane --- Certainty=0.3463(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9097> which encodes the amino acid sequence <SEQ ID 9098>. Analysis of this protein sequence reveals the following:

40

>>> May be a lipoprotein

45

----- Final Results -----  
 bacterial membrane --- Certainty= 0.000(Not Clear) < succ>  
 bacterial outside --- Certainty= 0.000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 66/251 (26%), Positives = 111/251 (43%), Gaps = 27/251 (10%)

50

Query: 23 PNSTSANQLSEIKSGVLKVGKQDVPNFGYYNAETNQYEGMEIDIAKKIAKSLGVKPVF 82  
 P+ + + IK+ GVLKV +YN + N+ G E+D+ K+I K L +K F  
 Sbjct: 34 PHQSQKSSWDTIKEKGVKLVATPGTYQPTSFYN-DNNELVGVEYEDMVKEIGKRLNIKVKF 92

55

Query: 83 VPTTAQTREPLMDNGQIDILIATYTTITPERKANYNISKAYYHDEIGFLVR---KNSHIK 138  
 V T +D+G++DI + + ITP+R+ YNIS Y + G +VR N K  
 Sbjct: 93 VETGFDQAFTSVDSGRVDISLNNFDITPKRQKKYNISTPYKYGVGGMIVRADGSSNIAKK 152

60

Query: 139 TIKELDGKHIGVAQGATTKVNLEKYAKEHKLKFSYAQLGSPPELAISLYANRI----- 191  
 + + GK A G +K A+L ++ + +Y N +  
 Sbjct: 153 DLSDWKGGKAAGASGTEYMKVAQKQG-----AELVTYDNTVDVYLNDVANGRTDF 203

Query: 192 --DAFSVDKSILSGYLSPHTTILKE---GFNTQYGIATSKQDKVLIPYVNKLLVSWEK 245



-463-

++ K + LS + + + +N E GI +K+D L ++ ++ K  
 Sbjct: 204 IPNDYPAQKLFDVYMLSQNPNLNVKMSDVQYNPTEQGIVMNNKDDSLKKKIDAVIKDMIK 263

Query: 246 DGSLKHIYQKF 256

DGSLK I + +

Sbjct: 264 DGSLKKISETY 274

SEQ ID 1188 (GBS136) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 5; MW 29.9kDa).

10 The GBS136-His fusion product was purified (Figure 200, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 284), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 15 Example 366

A DNA sequence (GBSx0397) was identified in *S.agalactiae* <SEQ ID 1191> which encodes the amino acid sequence <SEQ ID 1192>. This protein is predicted to be integral membrane. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.34	Transmembrane	32 - 48 ( 27 - 55)
INTEGRAL	Likelihood = -5.04	Transmembrane	200 - 216 ( 196 - 219)
INTEGRAL	Likelihood = -3.13	Transmembrane	93 - 109 ( 93 - 113)
INTEGRAL	Likelihood = -2.02	Transmembrane	74 - 90 ( 74 - 92)

----- Final Results -----

bacterial membrane --- Certainty=0.4736(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB73177 GB:AL139076 putative ABC-type amino-acid transporter  
 permease protein [Campylobacter jejuni]

Identities = 112/226 (49%), Positives = 160/226 (70%), Gaps = 3/226 (1%)

Query: 5 NISPF AISRWGAFFNHF DLFFKGF LYT LGISFGALL LAL ILGILSGGLSTSKSKV GKLIS 64

+ISPFA+ ++ ++ D F GF+YTL +S ALL+A I G + G ++TS+ K+ + +

Sbjct: 25 SISPF AVWKFLDALDNKDAFINGFIYTLEVSILALLIATIFGTIGGVMATSRFKIIRAYT 84

Query: 65 RIYVEVFQNTPLLVMVFVYGLAIISNGHVMISAFFTAVLCVGLYHGAYISEVIRSGIE 124

RIYVE+EQN PL++Q+ F++Y L ++ + + F VL VG YHGAY+SEV+RSGI

Sbjct: 85 RIYVELFQNVPLVIQIFFLFYALPVLG---IRLDIFTIGVLGVGAYHGAYVSEVVRSGIL 141

Query: 125 AVPKGQTEAALAQGFTANQTMQLIILPQAVRTILPPMTNQVNNLIKNTSTVAIISGADIM 184

AVP+GQ EA+ +QGFT Q M+ II+PQ +R ILPPMTNQ+VNLIKNTS + I+ GA++M

Sbjct: 142 AVPRGQFEASASQGFTYIQQMRYIIVPQTIRIILPPMTNQMVNNLIKNTSVLLIVGGAELM 201

Query: 185 FVAKAWAYDTTNYIPAFAGAAIFYFVICFPLASWARKQEELNKKTY 230

A ++A D NY PA+ AA+ YF+IC+PLA +A+ E KK +

Sbjct: 202 HSADSYAADYGNYAPAYIFAALVLYFIICYPLAYFAKAYENKLKKAH 247

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1193> which encodes the amino acid sequence <SEQ ID 1194>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

-464-

INTEGRAL	Likelihood = -6.26	Transmembrane	307 - 323 ( 303 - 327)
INTEGRAL	Likelihood = -5.89	Transmembrane	485 - 501 ( 479 - 502)
INTEGRAL	Likelihood = -1.12	Transmembrane	375 - 391 ( 375 - 391)

## 5 ----- Final Results -----

bacterial membrane --- Certainty=0.3506(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

## 10 The protein has homology with the following sequences in the databases:

&gt;GP:BAA17584 GB:D90907 glutamine-binding periplasmic protein

[Synechocystis sp.]

Identities = 146/532 (27%), Positives = 244/532 (45%), Gaps = 59/532 (11%)

15 Query: 6 YMKKLILSCLVALALLFGGMSRAQANQYLRVGMEEAYAPFNWTQDDASNGAVPIEGTSQY 65  
Y L L L+A+A+ + Q + V E + PF T E T Q  
Sbjct: 16 YYLLLLALGVLLAIAIPLLPFAFSQVSRQTIIVATEPTFPFPFEMTD-----EATGQL 65

20 Query: 66 ANGYDVQVAKKAKAMNKELLVVKTSWTGLIPALTSGKIDMIAAGMSPTKERRNEISFSN 125  
G+DV + + + +A + + + G+IPAL S + + + + T ER +SFS+  
Sbjct: 66 T-GFVDVLIQAIGEAQVTVDIQGYPFDDGIIPALQSNVTGAAISAITITPERAQSVSFSS 124

25 Query: 126 SSYTSQPVLVVTANGKYADATSLKDFSGAKVTAQQGVVHVNLLTQLKGAKLQTPMGDFSQ 185  
+ S VL + +LKD G ++ G + T + GAK+ T +  
Sbjct: 125 PYFKS--VLAIAVQDGNNTIKNLKDLGKRLAVAIGTTGAMVATNVPGAKV-TNFDSITS 181

30 Query: 186 MRQALTSVIDAYISERPEAMTAEAADSRLLKMITLKKGFAVAESDAAIAVGMKKNDDRMA 245  
Q L +G DA I++RP + A D+ L+ + + +E IA+ + +  
Sbjct: 182 ALQELVNGNADAVINDRPFVLLYA-IKDAGLRNVKISADVG-SEDYVGIAMPLAPPGE--- 236

35 Query: 246 TVNQVLEGFSGQTDRLMALMDMVTQKQVEKKAEDAKASFLGQMWAIKGN----- 294  
+NQ E +Q ++++ EK + FL + G  
Sbjct: 237 -INQTRVFLNQ-GLFQIENGTYNAIYEKWFGEKNPPFLPLVAPSLVGKVGTAQSLTERS 294

40 Query: 343 LFGWLLTIYIEIFRGTPMIVQSMVIYYGTAQAF-----GISIDRTLAAIFIVSINTGAYM 397  
+ IY+E FRGTPM+VQ +IY+G F GI+IDR AAI +S+N AY+  
Sbjct: 345 PLQLIFRIYVEFFRGTPMLVQLFIIYFGLPALFKEIGLGITIDRFPAIIALSINVAAYL 404

45 Query: 398 SEIVRGGIFAVDKGQFKAATALGFTHGQTMRKIVLPQVVRNLPATGNEFVINIKDTSVL 457  
+EI+RGGI ++D+GQ++A +LG + QTM++++ PQ R ILP GNEF+ IKDTS+  
Sbjct: 405 AEIIRGGIQSIDQGQWEACESLGMSPWQTMKEVIFPQAFRRILPPLGNEFITLIKDTSLT 464

50 Query: 458 NVISVVVELYFSGNTVATQTYQYFQTFTIIAIIYFVLTFTVTRILRYIERRFD 509  
VI EL+ G + TY+ F+ + +A++Y +LT + + +++E D  
Sbjct: 465 AVIGFQELFREGQLIVATTYRAFEVYIAVALVYLLLTITISSFVFKWLENYMD 516

An alignment of the GAS and GBS proteins is shown below:

Identities = 82/210 (39%), Positives = 113/210 (53%), Gaps = 12/210 (5%)

55 Query: 14 WGAFFNHFDLFFKGFYTLGISFGALLLALILGILSGGLSTS---KSKVGKL-----I 63  
W F ++ F +G TL IS + L +G+L G T+ K KV L +  
Sbjct: 288 WAIFKGNWKQFLRGTMILLISMVGTITGLFIGLLIGIFRTAPKAKHKVAALGQKLFGL 347

60 Query: 64 SRIYVEVFQNTPLLQMVVYVYGLAISNGHVMISAFFTAVLCVGLYHGAYISEVIRSGI 123  
IY+E+F+ TP++VQ + +YVG A + I A+ V + GAY+SE++R GI  
Sbjct: 348 LTIYIEIFRGTPMIVQSMVIYYGTAQAFG--ISIDRTLAAIFIVSINTGAYMSEIVRGGI 405

65 Query: 124 EAVPKGQTEAALAQGFTANQTMQLIILPQAVRTILPPMTNQVNNLIKNTSTVAIISGADI 183  
AV KGQ +AA A GFT QTM+ I+LPQ VR ILP N+ V IK+TS + +IS ++  
Sbjct: 406 FAVDKGQFKAATALGFTHGQTMRKIVLPQVVRNLPATGNEFVINIKDTSVLNVISVVEL 465

Query: 184 MFVAKAWAYDTTNYIPAFAGAAIFYFVICF 213

-465-

F A T Y F AI YFV+ F  
 Sbjct: 466 YFSGNTVATQTYQYFQTFTIIAIYFVLTF 495

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 367**

A DNA sequence (GBSx0398) was identified in *S.agalactiae* <SEQ ID 1195> which encodes the amino acid sequence <SEQ ID 1196>. This protein is predicted to be amino acid ABC transporter, permease protein. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -6.95 Transmembrane 25 - 41 ( 16 - 42)  
 INTEGRAL Likelihood = -3.61 Transmembrane 66 - 82 ( 65 - 86)  
 INTEGRAL Likelihood = -2.44 Transmembrane 184 - 200 ( 182 - 201)  
 INTEGRAL Likelihood = -0.59 Transmembrane 119 - 135 ( 119 - 135)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.3781(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14704 GB:Z99118 glutamine ABC transporter (integral membrane protein) [Bacillus subtilis]  
 Identities = 84/206 (40%), Positives = 129/206 (61%), Gaps = 6/206 (2%)  
 Query: 10 ILFLLQGFGTLTYISFISILLSMFFGTLTLLAIMRNSKNPIWKLIASIYIEFVRNVPNLLWI 69  
 + FL GF +TLY++FISI+LS FFG + +R +K P+ + ++ +E +RN+P LL I  
 Sbjct: 12 LAFLWDGFLVTLTYVAFISIIISFFGLIAGTLRYAKVPVLSQLIAVLVETIRNLPLLLLII 71  
 Query: 70 FIIFLVF-----QMKSVSAGITSFTIFTSAALAEIIRGGLNGVDKGQTEAGLSQGFTYLQ 124  
 F F +++ +A IT+ TIF SA L+EIIR GL +DKGQ EA S G +Y Q  
 Sbjct: 72 FFTFFALPEIGIKLEITAAAITALTIFESAMLSEIIRSGLSIDKGQIEAARSSGLSYTQ 131  
 Query: 125 VFIIIIIFPQAFRKMLPAIISQFVTVIKDTSLLYSVIAIQEIFGKSQILMGRYFEAGQVFT 184  
 I+ PQA R+M+P I+SQF++++KDTSL VIA+ E+ +QI+ G+ + F  
 Sbjct: 132 TLFFIVMPQALRRMVPIVSQFISLLKDTSLAV-VIALPELIHNAQIINGQSADGSYFFP 190  
 Query: 185 LYAIIITAVYFITNFIISFSRKLKR 210  
 ++ + +YF N+ +S +R+L R  
 Sbjct: 191 IFLAALMYFAVNYSLAARRLEVR 216

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1197> which encodes the amino acid sequence <SEQ ID 1198>. Analysis of this protein sequence reveals the following:

Possible site: 20  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -10.51 Transmembrane 529 - 545 ( 517 - 551)  
 INTEGRAL Likelihood = -10.30 Transmembrane 697 - 713 ( 693 - 719)  
 INTEGRAL Likelihood = -4.41 Transmembrane 560 - 576 ( 555 - 585)  
 INTEGRAL Likelihood = -0.32 Transmembrane 662 - 678 ( 662 - 678)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5203(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAA17584 GB:D90907 glutamine-binding periplasmic protein

-466-

[Synechocystis sp.]

Identities = 153/475 (32%), Positives = 251/475 (52%), Gaps = 27/475 (5%)

5 Query: 273 IVSDSSFPFEPQN-GKGKYVGIDIELIKAIKQGGFKIEIANPGFDAALNAVQSSQADG 331  
+ ++ +F PFE + G+ G D++LI+AI + ++I FD + A+QS+  
Sbjct: 46 VATEPTFPFEMTDEATGQLTGFDVDLIQAIGEAAQVTVDIQGYPFDDGIIPALQSNVTGA 105

10 Query: 332 VIAGATITDARKAIFDFSDPYYSNIIILAVKAGKN-IKNYEDLDRKTVGAKNGTSSYSWL 390  
I+ TIT R FS PY+ S + +AV+ G + IKN +DL+ K + GT+ + +  
Sbjct: 106 AISAITITPERAQSVSFSSPYFKSVLAIAVQDGNNTIKNLKDLGKRLAVAIGTTG-AMV 164

15 Query: 391 KENAPKYGYNVKAFDDGSSMYDSLNSGSVDAIMDDEAVLKYAIISQ--RRFETPLEGIST 448  
N P G V FD +S L +G+ DA+++D VL YAI R + + S  
Sbjct: 165 ATNVP--GAKVTNFDSITSALQELVNGNADAVINDREVLLYAIKDAGLRNVKISADVGSE 222

20 Query: 449 GEVGFVAVKKGITNPELI---EMFNGLAALKKSGQYDDIIDKYLDKKA-----ATPSEKG 500  
G A+ E+ E+ N GL + ++G Y+ I +K+ K PS G  
Sbjct: 223 DYYGIAMPLAPPGEINQTRVNLQGLFQIIENGTYNAIYEKWFGEKNPPFLPLVAPSLVG 282

25 Query: 501 -----ADESTISGLLSNNYKQLLAGLGTTLSTLISFAIAIIIGIIFGMMAVSP 549  
+ + L ++ L G T+ LT S +I G + +S  
Sbjct: 283 KVGTAQSLTERSQANPNDNFLITLFRNLFKGSILTVLLTAFSVFFGLIGGTGVAIALISD 342

30 Query: 550 TKSLRLISTVFDVVRGIPLMIVAIFWGVNPLIESMTGHQSPINDFLAATIALSLNGG 609  
K L+LI ++V+ RG P+++ I++G+P L + + G I+ F AA IALSIN  
Sbjct: 343 IKPLQLIFRIYVEFRGTPLVLQFLFIYFGLPALFKEI-GLGITIDRFPAIIALSINVA 401

35 Query: 610 AYIAEIVRGGIEAVPAGQMEASRSLGLSYGTTMRKVILPQAVKLMPLNFQFVVISLKDT 669  
AY+AEI+RGGI+++ GQ EA SLG+S TM++VI PQA + +LP N+F+ +KDT  
Sbjct: 402 AXLAEIIRGGIQSIDQGWEEACSLGMSWPQTMKEVIFPQAFRRILPPLGNEFITLIKDT 461

40 Query: 670 TIVSAIGLVELFQTKIIIRNYQSFRMYAILAIYILIMILLTRAKRLEKRLN 724  
++ + IG ELF+ G++I+A Y++F +Y +A++YL++ + + + K LE ++  
Sbjct: 462 SLTAVIGFQELFREGQLIVATTYRAFEVYIAVALVYLLTTISSFVKWLENYMD 516

45 Identities = 68/247 (27%), Positives = 106/247 (42%), Gaps = 11/247 (4%)

50 Query: 7 VLLLAIMSIFLTCNIASEAETIAIVSDTAYAPFEFKD--SDQIYKGIDVDIINEVAKRQSW 64  
VLL + + + S +TI + ++ + PFE D + Q+ G DVD+I + +  
Sbjct: 24 VLLAIAIPLLPAFSQVSRQTIIVATEPTFPFEMTDEATGQL-TGFDVDLIQAIGEAAQV 82

55 Query: 65 DFSMSFPFGDAAVNAVQSQASALMAGTTITNARKKVHFHSEPPYDTKIVIAIRKAN-AI 123  
+ FD + A+QS A ++ TIT R + FS PY+ + + IA + N I  
Sbjct: 83 TVDIQGYPFDDGIIPALQSNVTGAIAISAITITPERAQSVSFSSPYFKSVLAIAVQDGNNTI 142

60 Query: 124 KKYSDLKGGKTVGVKNGTAAQAFLLNNYKKKYDYTVKTFDTGDLMYNSLSAGSIAAVMDDEA 183  
K DL+GK + V GT N V FD+ L G+ AV++D  
Sbjct: 143 KNLKDLGKRLAVAIGTTGAMVATNVP--GAKVTNFDSITSALQELVNGNADAVINDRP 199

65 Query: 184 VIQYAIS----QNQDIAINMKGEPIGSFGFVAVKKGSGYDYLVDNFNTALKAMKADGTYQA 239  
V+ YAI +N I+ ++ E + + N L + +GTY A  
Sbjct: 200 VLLYAIKDAGLRNVKISADVGSEDDYYGIAMPLAPPGEINQTRVNLQGLFQIIENGTYNA 259

Query: 240 IMTKWL 246  
I KW G  
Sbjct: 260 IYKWF 266

An alignment of the GAS and GBS proteins is shown below:

Identities = 68/210 (32%), Positives = 113/210 (53%), Gaps = 16/210 (7%)

60 Query: 13 LLQCGFLTYISFISILLSMFFGTLLAIMRNSKNPIWKLIASIIYIEFVRNVNLLWIFII 72  
LL G G TL ++ IS +++ G + +M S +LI+++++ VR +P ++ I  
Sbjct: 517 LLAGLGTTLSTLISFAIAIIIGIIFGMMAVSPTKSLRLISTVFDVVRGIPLMIVAIFI 576

65 Query: 73 F-----LVFQMKSVSAGITSFTIFT-----SAALAEIIRGGLNGVDKGQTEAGLSQGF 120  
F L+ M + I F T A +AEI+RGG+ V GQ EA S G  
Sbjct: 577 FWGVPNPLIESMTGHQSPINDFLAATIALSLNGGAYIAEIVRGGIEAVPAGQMEASRSLGL 636

-467-

Query: 121 TYLQVFIIIIFFPQA FRKMLPAIISQFVTVIKDTSLLYSVIAIQEIFGKSQILMG RYFEAG 180  
 +Y +I PQA + MLP I+QFV +KDT+++ S I + E+F +I++ R +  
 Sbjct: 637 SYGTTMRKVLPQAVKLM L PNFNQFVISLKD T TIV-SAIGLVELFQTGKII IARNY--- 692

5 Query: 181 QVFTLYAIITAVYFITNFISSFSRKL SKR 210  
 Q F +YAI+ +Y I +++ +++L KR  
 Sbjct: 693 QSFRMYAILAI IY LIMIILL LTR LAKRLEKR 722

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 10 vaccines or diagnostics.

### Example 368

A DNA sequence (GBSx0399) was identified in *S. galactiae* <SEQ ID 1199> which encodes the amino acid sequence <SEQ ID 1200>. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -12.21 Transmembrane 7 - 23 ( 1 - 30)

----- Final Results -----  
 bacterial membrane --- Certainty=0.5883(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP: BAB04094 GB: AP001508 unknown conserved protein in B. subtilis  
 [Bacillus halodurans]  
 Identities = 43/157 (27%), Positives = 83/157 (52%), Gaps = 9/157 (5%)

Query: 26 YQSQFQKTTNQALAIAYKDAKVAKK--DVIHQKIDKEFENFRGSYEIEFNTKSAEYSYHV 83  
 +Q++ N+ L +A ++ + + + +K+ +N R YEIE EY + +  
 30 Sbjct: 38 HQAESVSADNEGLTLAEASDIALERAGNGVVTAEKDRDNGRVVYEIEVKNDDEYDFKI 97

Query: 84 DVKTGQILERDMDNNGFSKSTSQSSSSSQKSHKISQEEAKKIAFKDANIEESEVSNLKI 143  
 D +TG+IL+ + SK SSS ++ IS +EAK+IA K+ + ++ +++  
 Sbjct: 98 DQQTGEILKEKQEQRKSGKPREGHSSSKGSEA-VISMDEAKEIALKEVS---GKIDDIEL 153

35 Query: 144 KEEIENGKSVYDIDF-VDLKNKNEVDYQIDAETGKII 179  
 E ENG VY+++ D + ++V +DA TG ++  
 Sbjct: 154 --ERENGSLVYEVEIESDHYDDDDVTYVYVDAMTGNVL 188

40 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1201> which encodes the amino acid sequence <SEQ ID 1202>. Analysis of this protein sequence reveals the following:

Possible site: 57  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -5.15 Transmembrane 42 - 58 ( 41 - 60)

----- Final Results -----  
 bacterial membrane --- Certainty=0.3060(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

Identities = 37/96 (38%), Positives = 63/96 (65%), Gaps = 5/96 (5%)

55 Query: 94 DMDNNGFSKSTSQSSSSSQKSHKISQEEAKKIAFKDANIEESEVSNLKIKEE IENGKSV 153  
 DMD+ +Q +S + K K+S+++AK IA KDA++ E++ L + ++ E+GK+V  
 Sbjct: 59 DMDDKD-DHMDNQPKTSQTSKKVKLS EDKAKSIALKDASVTEADAQMLS VTDQDNEDGKAV 117

Query: 154 YDIDFVDLKNKN-EVDYQIDAETGKLIERSRDHMD 188  
Y+I+F +NK+ E Y IDA +G I+E+S + +ND  
Sbjct: 118 YEIEF---QNKDQEYSYTIDANSGDIVEKSSEPIND 150  
Identities = 23/62 (37%), Positives = 37/62 (59%)

A related GBS gene <SEQ ID 8563> and protein <SEQ ID 8564> were also identified. Analysis of this protein sequence reveals the following:

\*\*\* Reasoning Step: 3

```

----- Final Results -----
      bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

26.1/59.2% over 140aa  
 Bacillus subtilis  
 EGAD|107494| hypothetical protein Insert characterized  
 GP|2632048|emb|CAA05607.1|AJ002571 YkoJ Insert characterized  
 GP|2633682|emb|CAB13185.1||Z99110 similar to hypothetical proteins from B. subtilis  
 Insert characterized  
 PIR|F69859|F69859 conserved hypothetical protein ykoJ - Insert characterized

```

ORF00925(379 - 852 of 1164)
EGAD|107494|BS1329(29 - 169 of 170) hypothetical protein {Bacillus subtilis}
GP|2632048|emb|CAA05607.1|AJ002571 YkoJ {Bacillus subtilis}
GP|2633682|emb|CAB13185.1|Z99110 similar to hypothetical proteins from B. subtilis
{Bacillus subtilis} PIR|F69859|F69859 conserved hypothetical protein ykoJ - Bacillus
subtilis
%Match = 6.2
%Identity = 26.1 %Similarity = 59.2
Matches = 37 Mismatches = 52 Conservative Sub.s = 47

```

```

297      327      357      387      417      447      468      498
NIIE**KEGCCMIKKNKVFLEVLVLVLVILEGGVLFYQSQFQKTTNQAALAIAYKDAKVAKKDVIH---QKIDKEFENFRG
      |      :|      |      :: :: :: |: ::|: | |:      : ||::| :
      MLKKKWMVGLLAGCLAAGGFSYNAFATENNENRQASSKTDLTEQEAELIAKTVDGTVEDIDRDLYNGKE
      10      20      30      40      50      60      70

528      558      588      618      648      672      702      732
SYEIEFNTKSAEYSYHVDVKTGQILERDMDNNGFSKSTSQSSSSSSQKSHK--ISQEEAKKIAFKDANIEESEVSNLKIK
||:|      : :|      :||: | | |      :      :|: : |::||:|::|:|
VYEVEIEKEGEDYDVYVDIHTKQALNDPL-----KEKAEQVAITKEEAELIALKQTG---GTVTESKLD
      90      100      110      120      130

```

-469-

```

762      792      822      852      882      912      942      972
EEIENGKSVYDIDFVLDLKNKNEVDYQIDAETGKIIERSRDHMD*FK*DIKKRRSKRPSF*LLSSLLPTF*KFT*KT*DD
|:  :|  :|:::  :  |  |:::|  |:  |:|:::  |
ED--DGAYIYEME-IQTKQGTETETEFESAKDGRIIKQEIDD
5      140      150      160      170

```

SEQ ID 8564 (GBS37) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 4; MW 22kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 10; MW 47kDa).

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 369

A DNA sequence (GBSx0400) was identified in *S.agalactiae* <SEQ ID 1203> which encodes the amino acid sequence <SEQ ID 1204>. Analysis of this protein sequence reveals the following:

```

15  Possible site: 53
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.1499(Affirmative) < succ>
20  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9709> which encodes amino acid sequence <SEQ ID 9710> was also identified.

- 25 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1205> which encodes the amino acid sequence <SEQ ID 1206>. Analysis of this protein sequence reveals the following:

```

    Possible site: 42
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2808(Affirmative) < succ>
30  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

35 An alignment of the GAS and GBS proteins is shown below:

Identities = 128/297 (43%), Positives = 180/297 (60%), Gaps = 9/297 (3%)

```

40  Query: 54  IDDIKVGSPIFKYFWT-SLSLQAPLKALEFVLEQAKMPTELSGELSETQYLVAQFSDELA 112
      I D ++GSP F   W      Q+  + L F+L+  +MP ELSG+L ETQ L+ +F   L
      Sbjct: 46  IIDNRLGSPTFWIWIPIEKENQSAKQLLTFLLDLVEMPFELSGQLHETQTLTTRFHPSLL 105

      Query: 113 PHDDFWIALSQVIYDSFPGNSLAEDTVLNRKLHQFRYLISQQQAQYVRRYFKDVGMTDRD 172
      P   FW L+ ++  +FPG +L++  L ++LHQFRY+ISSQQAQ +R ++K + MTD
45  Sbjct: 106 PDHMFVKELASLVDQAFPGKTLAQAGELEKRLHQFRYVISSQQAQSIRNHYKMIEMTDAQ 165

      Query: 173 ALVNYL-----SCL-REPDSIAYYESARLHNKRRRNGEIFGFPDDEPVINSKLLISFHT 226
      AL  +L      CL R+      +SARLHNK R      FP E   N K+L+ FHT 225
      Sbjct: 166 ALALFLRSKKGPCLRQAPDYTLMD SARLHNKLRFDENKVIFFPSQEVSYNIKVLLWFHT 225

50  Query: 227 FIIDDKGNFLNEIDA EVITRNGIINGASFNYAFKNNTRHKELDVDPVK-LDPKFRNDMTR 285
      F +D  G FLNE+DAEV+T  GI+NGASFNY  +  RH  +LDVDP+  DP+FR D  +
      Sbjct: 226 FTL DSTGFFLNEVDAEVVTEKGIVNGASFNYG-TDGPRHWDLDVDPISHHDPQFRDRTLK 284

```

-470-

Query: 286 GYRSPNLSRRKWFFFKEDYDCSYFNKKGYAFGRSAKQSVDKQVKYLKKAQVQKMR 342  
 G+RSP R+WF +++D+ SYFN KG +A+ +S+ V K K K+ + ++  
 Sbjct: 285 GFRSPKRVFRQWFRAQKDDFMFSYFNAKGLFAYHNKSSFARVKKSAKQFKRQIHPIK 341

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 370

- A DNA sequence (GBSx0401) was identified in *S.agalactiae* <SEQ ID 1207> which encodes the amino acid sequence <SEQ ID 1208>. This protein is predicted to be similar to two-component response regulator [YcbM] (ompr-likeprotei). Analysis of this protein sequence reveals the following:

Possible site: 31  
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3129(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

- 20 >GP:CAA55264 GB:X78502 gtcR [Brevibacillus brevis]  
 Identities = 99/228 (43%), Positives = 149/228 (64%), Gaps = 3/228 (1%)
- Query: 2 RTVLVVQGDDETIELLRSYLEGALYKVVMSDGEAFSLFQHQIDLAIIDITLPKIDGY 61  
 +T+L+ + E IELL+ +LE Y+++ A DGE+A++ +QH +DLAII DI +P +DG+  
 25 Sbjct: 3 KTLIADDEPEIIELLKLFLERESYRIIEAYDGEQAWNYIRQHPVDLAIIDIMPALDGF 62
- Query: 62 ELTRLIRQDSQIPIIMLAAKTDMRILGLNIGADDFITKPFNSLEVLARINSQLRRYYE 121  
 +L + + + ++P+I+L+AK D D+ILGL +GADDFI+KPFN LE +ARI +QLRR +E  
 30 Sbjct: 63 QLIKRLTNEYKLPVILSAKNRSDKILGLGLGADDFISKPFNPLEAVARIQAQLRRAFE 122
- Query: 122 FNSLAKP--KNQFIKIGELELDEEHVELTKNGKHIKLTATEFKILHILMS-SPGRIYTKT 178  
 FN + Q +G L L + + + +T E+++L+ M S I+TK  
 35 Sbjct: 123 FNEPEEKAISTQSTTVGRLTLHTACVVYRGDETYSVTPLEYRLNITFMQCSRTSIFTKQ 182
- Query: 179 QLYEKINGRYLEGDETTIMVHISNIRDKIEDDSKYPKYIKTLRGVGYK 226  
 QL+E+ D+ TIMV IS +RDKIED + P YIKT+RG+GYK  
 Sbjct: 183 QLFEQAWSETYWEDDNTIMVQISRLRDKIEDQPRQPVYIKTVRGLGYK 230

There is also homology to SEQ ID 1182:

- 40 Identities = 87/230 (37%), Positives = 144/230 (61%), Gaps = 5/230 (2%)
- Query: 1 MRTVLVVQGDDETIELLRSYLEGALYKVVMSDGEAFSLFQHQIDLAIIDITLPKIDG 60  
 M+ +L+V + +++++ L Y +V A DG EA ++F++ + DL I+D+ LP++DG  
 45 Sbjct: 1 MKKILIVDDEKPISDIIKFNLTKEGYDIVTAFDGREAVTIFEEKPDLIILDLMLPELDG 60
- Query: 61 YELTRLIRQDSQIPIIMLAAKTDMRILGLNIGADDFITKPFNSLEVLARINSQLRRYY 120  
 E+ + IR+ S +PII ML+AK ++ D+++GL IGADD++TKPF++ E+LAR+ + LRR  
 Sbjct: 61 LEVAKEIRKTSHPVPII MLSAKDSEFDKVIGLEIGADDYVTKPFNSRELLARVKAHLRTE 120
- 50 Query: 121 EFNSLAKPKN-----QFIKIGELELDEEHVELTKNGKHIKLTATEFKILHILMSSPGRIY 175  
 + +N Q + IG L++ + K+G+ ++LT EF++LH L + G++  
 Sbjct: 121 TIETAVAEENASSGTQELTIGNLQILPDFAVAKKHGQVELTHREFELLHHLANHMQVM 180
- Query: 176 TKTQLYEKINGRYLEGDETTIMVHISNIRDKIEDDSKYPKYIKTLRGVGY 225  
 T+ L E + G GD T+ V + +R+KIED P+YI T RGVGY  
 55 Sbjct: 181 TREHLLLEIVWGYDYFGDVRTVDVTVRRLREKIEDTPSRPEYILTRRGVGY 230

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.



-471-

**Example 371**

A DNA sequence (GBSx0402) was identified in *S.agalactiae* <SEQ ID 1209> which encodes the amino acid sequence <SEQ ID 1210>. This protein is predicted to be threonyl-tRNA synthetase 1 (thrS). Analysis of this protein sequence reveals the following:

```

5   Possible site: 32
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2353(Affirmative) < succ>
10   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15   >GP:BAB06860 GB:AP001517 threonyl-tRNA synthetase 1 [Bacillus halodurans]
    Identities = 413/638 (64%), Positives = 506/638 (78%), Gaps = 7/638 (1%)

    Query: 1   MIKITFPDGAIREFESGITTFEIAQSISNSLAKKALAGKFNGLIDTTTRAIEEDGSIEIV 60
    Sbjct: 4   MINITFPDGAIVKEFPKGTTTAEIAGSISPLKKKALAGMLDGTLLDLNTPIEQDGTITIV 63

20   Query: 61   TPDHEDALGVLRRHSAHLFAQAARKLFPPD--LCLGVGPAIQDGFYYDTDNKSGQISNDDL 118
    TP+ ++AL VLRHS AH+ AQA KRLF D + LGVGP I+ GFYYD D ++ +DL
    Sbjct: 64   TPESDEALEVLRHSTAHVMAQALKRFLKDRNVKLGVPVIEGGFYDVMDES-LTPEDL 122

25   Query: 119  PRIIEEMKKIVKENHPCIREEISKEEALFLFKD--DPYKVELISEHAEDG-LTVYRQGEF 175
    P+IE+EMKKI+ EN P R +S+EEAL +++ DPYK+ELI++ ED +T+Y QGEF
    Sbjct: 123  PKIEKEMKKIIGENLPIERVVVSREEALARYEEVGDPYKIELINDLPEDETITITYEQGEF 182

30   Query: 176  VDLCRGPHVPSTGRIQVFHLLNVAGAYWRGNSDNAMQRVYGTAWFDKKDLKAYLKRREE 235
    DLRCG HVPSTG+++ F LLN+AGAYWRG+S N M+QR+YGTA+F K DL +L+ EE
    Sbjct: 183  FDLRCGVHVPSTGKLKEFKLLNLAGAYWRGDSSNKMLQRIYGTAFKKADLDEHLRLLEE 242

    Query: 236  AKERDHRKLGKELDLFMVNPEVGQGLPFWLPNGATIRRELEERYIVDKEIASGYQHVVYTP 295
    AKERDHRKLGKEL +F ++ +VGQGLP WLP GATIRR +ERYIVDKE GYQHVVYTP
35   Sbjct: 243  AKERDHRKLGKELGIFALSQKVGQGLPLWLPKGATIRRIERYIVDKEEKLGYQHVVYTPV 302

    Query: 296  MASVEFYKTSGHWDHYREDMFPTMDMGDGEFVLRPMNCPHHIEVYKHHVHSYRELPIRI 355
    +AS E YKTSGHWDHY++DMFPTM+M + EE VLREPMNCPHH+ VYK + SYR+LP+RI
    Sbjct: 303  LASSELYKTSGHWDHYKDDMFPTMEM-ENEELVLRPMNCPHHMVYKTEMRSYRQLPLRI 361

40   Query: 356  AELGMMHRYEKSGALTGLQRVREMTLNDAHIFVTPEQIKDEFKALNLIABITYEDFNLT 415
    AELG+MHRYE SGA++GLQVR MTLNDAHIF P+QIKDEF++ + LI +YEDF L +
    Sbjct: 362  AELGLMHRYESGAVSGLQRVGMTLNDAHIFCRPDQIKDEFVVRVRLIQAVYEDFGLKN 421

45   Query: 416  YRFRLSYRDPEDKHXYNDNEMWENQAAMLEAMDDFGLDYFEAEGEAAFYGPKLDIQVK 475
    Y FRLSYRDPEDK KY+D+D MW AQ MLKEAMD+ L+YFEAEGEAAFYGPKLD+QV+
    Sbjct: 422  YSFRLSYRDPEDKEKYFDDDNMWNKAQGMLKEAMDELELEYFEAEGEAAFYGPKLDVQVR 481

50   Query: 476  TALGNEETLSTIQLDFLLPERFDLKYIGADGEEHRPIMIHRGGISTMERFTAILIETKYG 535
    TALG +ETLST+QLDFLLPERFDL Y+G DG+ HRP+++HRG +STMERF A L+E YKG
    Sbjct: 482  TALGKDETSTVQLDFLLPERFDLTVVGEDGQPHRPVVVHRGVVSTMERFVAFLLEEYKG 541

    Query: 536  AFPTWLAPQQVSVIPISNEAHIDYAEVARVLKDRGIRAEVDDRNEKMQYKIRAAQTQKI 595
    AFPTWLAP QV VIP+S EAH++YA V L+ GIR E+D+R+EK+ YKIR AQ QKI
55   Sbjct: 542  AFPTWLAPVQVQVIPVSPEAHLEYAKNVQETLQQAGIRVEIDERDEKIGYKIREAQMQKI 601

    Query: 596  PYQLIVGDKEEMEEKAVNVRRYSGKATETKSIEEFVESI 633
    PY L++GDKE+E VNVR+YG K + + ++EFV +
    Sbjct: 602  PYMLVLGDKEVEANGVNVRKYGEKDSSSMGLDEFVRHV 639

60

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1211> which encodes the amino acid sequence <SEQ ID 1212>. Analysis of this protein sequence reveals the following:

-472-

Possible site: 32

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

5                   bacterial cytoplasm --- Certainty=0.2566(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10       Identities = 564/644 (87%), Positives = 608/644 (93%)

Query: 1   MIKITFPDGAIREFESGITTFFETIAQSISNSLAKKALAGKFNQGLIDTTTAAIEEDGSIEIV 60  
           MIKITFPDGA+REFESG+TTF+IA+SIS SLAKKALAGKFN QGLIDTTTAAIEEDGSIEIV  
 Sbjct: 1   MIKITFPDGAIREFESGVTTFDIAESISKSLAKKALAGKFNQGLIDTTTAAIEEDGSIEIV 60

15       Query: 61   TPDHEDALGVLRHSAAHLFAQAARKLFPDLCLGVGPAIQDGFYYDTDNKSGQISNDDLPR 120  
           TPDH+DA VLRHSAAHLFAQAARKLFP+L LGVGPAL +GFYYDTDN QQISN+DLPR  
 Sbjct: 61   TPDHKDAYEVLHSAAHLFAQAARKLFPNLHLGVGPAIAEGFYDTDNNAEQISNDDLPR 120

20       Query: 121   IEEEMKKIVKENHPCIREEISKEEALFVKDDPYKVELISEHAEDGLTVYRQGEFVDLCR 180  
           IE EM+KIV EN+PCIREE++KEEALFVKDDPYKVELI+EHA GLTVYRQGEFVDLCR  
 Sbjct: 121   IEAEMQKIVTENYPCIREEVTKEEALFVKDDPYKVELINEHAGAGLTVYRQGEFVDLCR 180

25       Query: 181   GPHVPSTGRIQVFHLLNVAGAYWRGNSDNMMQRVYGTAWFDKDLKAYLKRREREKERD 240  
           GPHVPSTGRIQVFHLLNVAGAYWRGNSDN MMQR+YGTAWFDKDLKAYL R EAKERD  
 Sbjct: 181   GPHVPSTGRIQVFHLLNVAGAYWRGNSDNMMQRIYGTAWFDKDLKAYLTRLEAKERD 240

30       Query: 241   HRKLGKELDLFMVNPEVGQGLPFWLPNGATIRRELERYIVDKELASGYQHVVYTPPMASVE 300  
           HRKLGKELDLFM++ EVGQGLPFWLP+GATIRR LERYI DKE+ASGYQHVVYTPP+ASVE  
 Sbjct: 241   HRKLGKELDLFMISQEVGQGLPFWLPDGGATIRRTLERYITDKELASGYQHVVYTPPLASVE 300

35       Query: 301   FYKTSGHWDHYREDMFPTMDMGDEEFVLRPMNCPHHIEVYKHVHVSRELPIRIAEELGM 360  
           YKTSGHWDHY+EDMFP MDMGDGEEFVLRPMNCPHHI+VYK+HV SYRELPIRIAEELGM  
 Sbjct: 301   LYKTSGHWDHYQEDMFPMMDMGDEEFVLRPMNCPHHIQVYKNHVSRELPIRIAEELGM 360

40       Query: 361   MHRYEKSGALTGLQVRREMTLNDIAHIFVTPEQIKDEFKALNLIAEIVEDFNLTDIRFRL 420  
           MHRYEKSGAL+GLQVRREMTLND HIFVTPEQI++EF +AL LI ++Y DFNLTDIRFRL  
 Sbjct: 361   MHRYEKSGALSGLQVRREMTLNDGHIFVTPEQIQEEFQALQLIIDVYADFNLTDIRFRL 420

45       Query: 421   SYRDPEDKHYYDNDDEMWENAQAMLKEAMDDFGLDYFEAEGEAAFGPKLDIQVKTALGN 480  
           SYRDP D HKYYDNDDEMWENAQ+MLK A+D+ G+DYFEAEGEAAFGPKLDIQVKTALGN  
 Sbjct: 421   SYRDPNDTHKYYDNDDEMWENAQSMKAAALDEMGVDFEAEGEAAFGPKLDIQVKTALGN 480

50       Query: 481   EETLSTIQLDLFLPERFDLKYIGADGEEHRPIMIHRGGISTMERFTAILIETKGAFFPTW 540  
           EETLSTIQLDLFLPERFDLKYIGADGEEHRP+MIHRG ISTMERFTAILIETKGAFFPTW  
 Sbjct: 481   EETLSTIQLDLFLPERFDLKYIGADGEEHRPVMIHARGVISTMERFTAILIETKGAFFPTW 540

55       Query: 541   LAPQQVSVIPISNEAHIDYAWEVARVLKDRGIRAEVDDRNEKMQYKIRAAQTQKIPYQLI 600  
           LAP QV+VIPISNEAHIDYAWEVA+ L+DRG+RA+VDDRNEKMQYKIRA+QT KIPYQLI  
 Sbjct: 541   LAPHQVTVIPISNEAHIDYAWEVAKTLDRGVDRADVDDRNEKMQYKIRASQTSKIPYQLI 600

60       Query: 601   VGDKEMEEKAVNVRRYGSKATETKSIEEFVESILADIARKSRPD 644  
           VGDKEME+K+VNVRRYGSK T T+S+EEFVE+ILADIARKSRPD  
 Sbjct: 601   VGDKEMEDKSVNVRRYGSKTTHTESVEEFVENILADIARKSRPD 644

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 372**

60   A DNA sequence (GBSx0403) was identified in *S.agalactiae* <SEQ ID 1213> which encodes the amino acid sequence <SEQ ID 1214>. Analysis of this protein sequence reveals the following:

Possible site: 16

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

-473-

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1985(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA72250 GB:Y11463 ORF5 [Streptococcus pneumoniae]  
 Identities = 189/290 (65%), Positives = 234/290 (80%)

Query: 1 MRIGLFTDITYFPQVSGVSTSIRTLKEGLEKEGHEVYIFTTDRNVKRFEDPTIIRLPSVP 60  
 MRIGLFTDITYFPQVSGV+TSIRTLK LEK+GH V+IFTTDD+V R+ED IIR+PSVP  
 Sbjct: 1 MRIGLFTDITYFPQVSGVATSIRTLKTELEKQGHAVFIFTTDDKDVNRYEDWQIIRIPSPV 60

Query: 61 FISFTDRRVVYRGLISAYRIADKYELDIIHTQTTEFSLGGLGKLVAALRIPVVHTYHTQY 120  
 F +F DRR YRG A IAK Y+LDIIHTQTTEFSLGGLG +A+ L+IPV+HTYHTQY  
 Sbjct: 61 FFAFKDRRFAYRGFSKALEIAKQYQLDIIHTQTTEFSLGGLGIWIARELKIPVIHTYHTQY 120

Query: 121 EDYVGYIAKGKLIKPSMVKYIMRTYLSLDGVICPSRIVLNLLDGYGVKIPKQVIPTGIP 180  
 EDYV YIAKG LI+PSMVKY++R +L D+DGVICPS IV +LL Y VK+ K+VIPTGI  
 Sbjct: 121 EDYVHYIAKGLIRPSMVKYLVGRFLHDVDGVICPSEIVRDLLSDYKVKVEKRVPTGIE 180

Query: 181 VENYRREDISEETIKNLRTELGLADNDTMLSLSRSVSFEKNIQAAILMHLASAVVDENPHVK 240  
 + + R +I +E +K LR++LG+ D + LLSLSR+S+EKNIQA+L+ + V+ E VK  
 Sbjct: 181 LAKFERPEIKQENLKELRSKLGIDQGEKTLLSLSRISYEKNIQAVLVAFADVLKEEDKVK 240

Query: 241 LVIVGDGPYLSDLKELVHLSLELENSVIFTGMVEHSQVAIYYKACDFFISA 290  
 LV+ GDGPYL+DLKE +LE+++SVIFTGM+ S+ A+YYKA DFFISA  
 Sbjct: 241 LVVAGDGPYLNLDLKEQAQNLLEIQDSVIFTGMIAPSETALYYKAADFFISA 290

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1215> which encodes the amino acid sequence <SEQ ID 1216>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1074(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 309/444 (69%), Positives = 370/444 (82%)

Query: 1 MRIGLFTDITYFPQVSGVSTSIRTLKEGLEKEGHEVYIFTTDRNVKRFEDPTIIRLPSVP 60  
 MRIGLFTDITYFPQVSGV+TSIRTLK LEKEGHEVYIFTTDR+VKRFEDPTIIRLPSVP  
 Sbjct: 1 MRIGLFTDITYFPQVSGVATSIRTLKEELEKEGHEVYIFTTDRDVKRFEDPTIIRLPSVP 60

Query: 61 FISFTDRRVVYRGLISAYRIADKYELDIIHTQTTEFSLGGLGKLVAALRIPVVHTYHTQY 120  
 F+SFTDRRVVYRGLIS+Y+IAK Y LDIIHTQTTEFSLGGLGK++ KALRIPVVHTYHTQY  
 Sbjct: 61 FVSFTDRRVVYRGLISSYKIAKHYNLDIIHTQTTEFSLGGLGKMIGKALRIPVVHTYHTQY 120

Query: 121 EDYVGYIAKGKLIKPSMVKYIMRTYLSLDGVICPSRIVLNLLDGYGVKIPKQVIPTGIP 180  
 EDYV YIA GK+I+PSMVK ++R YL DLDGVICPSRIVLNLL+GY V IPK+VIPTGIP  
 Sbjct: 121 EDYVSYIANGKIIRPSMVKPLLRGYLKDLGVICPSRIVLNLLGEYEVTTIPKRVPTGIP 180

Query: 181 VENYRREDISEETIKNLRTELGLADNDTMLSLSRSVSFEKNIQAAILMHLASAVVDENPHVK 240  
 +E Y R+DI+ E + NL+ ELG+A ++TMLSLSR+S+EKNIQAI+ + A++ EN +K  
 Sbjct: 181 LEKYIRDDITAEVETNLKAEGLIAGDETMLSLSRISYEKNIQAIIINQMPAILAENAKIK 240

Query: 241 LVIVGDGPYLSDLKELVHLSLELENSVIFTGMVEHSQVAIYYKACDFFISASTSETQGLTY 300  
 L+IVG+GPYL DLK L LE++ V FTGMV H +VA+YYKACDFFISASTSETQGLTY  
 Sbjct: 241 LIIVGNPGLQDLKHLAMQLEVDKHTFTGMVPHDKVALYYKACDFFISASTSETQGLTY 300

Query: 301 IESLASGRPIIAQSNPYLDDVISDKMFGTLYKESDLADAILDAIAETPKMTQEAYEQKL 360

-474-

IESLASG PIIA NPYLDDV++DKMFGTLY E+DL DAI+DAI +TP M + +K  
 Sbjct: 301 IESLASGTPIIAHGNPYLDDVTDKMFGTLYAETDLTDAIIDAILKTPVMDKRLAKKR 360

Query: 361 YEISAENFSKSVYAFYLDLFLISQKASVKEKVSILTIGNKDSHSTLRFVRKAVYLPKKVFTF 420  
 YEISA++F KS+Y FYLD LI++ + +K+SL + + S+L+ V+ A++LPK+  
 Sbjct: 361 YEISAQHFGKSIYTFYLDTLIARNSKEAQKLSLYLNHSGKSSSLKLVQGAHLPKRAAKV 420

Query: 421 TGRASKKVVKAPKRRIRSDFLD 444  
 T S KVVKAP + + +I+DFLD  
 Sbjct: 421 TAITSVKVVKAPIKLVHAIKDFLD 444

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 373

A DNA sequence (GBSx0404) was identified in *S.agalactiae* <SEQ ID 1217> which encodes the amino acid sequence <SEQ ID 1218>. This protein is predicted to be lipopolysaccharide biosynthesis protein-related protein. Analysis of this protein sequence reveals the following:

Possible site: 61  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4076(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG19110 GB:AE005009 Vng0600c [Halobacterium sp. NRC-1]  
 Identities = 117/350 (33%), Positives = 178/350 (50%), Gaps = 29/350 (8%)

Query: 1 MKVLLYLEAEYLLKSGIGRAIKHQEKALQIAGIDYTTNPT----- 41  
 M+ L YLEA E L+ G+ A Q AL+ ++ P  
 Sbjct: 2 MRALNYLEAAEALR-GGMVTATNQRAALETVDVEVVETPWVAGDPVRSIGSLAAGGSCF 60

Query: 42 DDFDLVHMNTYGIIRSWLLMSKAKKTGKKVIMHGHSTEDFRNSFIGSNLVSPLFKWYLCR 101  
 FD+ H N G S + A++T +++H H T EDF SF GS+ ++P + YL  
 Sbjct: 61 TAFDVAHCNVLGPGSVAVARHARRTDTPVLVLAHLTREDFASFRGSSTIAPALEPYLRW 120

Query: 102 FYQKADAIITPTDYSKQLIKAYGIKKPIFVLSNGIDLRYQSEKKESAFRHYFHLKDD 161  
 FY +AD ++ P++Y+K +++AY + PI LSNG+DL Q E + R F L D  
 Sbjct: 121 FYSQADLVLCPEYTKDVLRAYPVDAPIRQLSNGVDLESMQGYESFRADTRARFDL--DG 178

Query: 162 KVMGAGLYFMRKGIDQFVEVAAKMPDIRFIWFGETNKWVPRKVRQIVTKQHPSNVTFA 221  
 VV G F RKG+ F E+ AK D F WFG ++ + P+NVTF  
 Sbjct: 179 TVVYAVGEVFERKGLTMFCCL-AKATDHEFAWFGPYDEGPQAGAATRKWVADPPANVTFT 237

Query: 222 GYIKGDVYEGAMSASDAFFFPRETEGIVVLEALASHQHVLRDIPVYHGWVTE-DSVE 280  
 GY++ A A D + FP++ E +GI VLEA+A + VVLRDIPV+ + T+ +  
 Sbjct: 238 GYMEDK--RAAFGAGDIYLFPAKVENQGIADVLEAMACGKPVVLRDIPVFREFFTDGEDCL 295

Query: 281 LATDVGDFVEKLDKVLGSKSDKIKEGYH--VAESRSIERIAHELASVYQ 327  
 + + + F + +D++ + + G + AES S++RI ELAS+Y+  
 Sbjct: 296 MCSTFEAFRDAIDRLADDPELRTRLGENARETAESHSLDRIGEELASIYE 345

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1219> which encodes the amino acid sequence <SEQ ID 1220>. Analysis of this protein sequence reveals the following:

Possible site: 61  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-475-

bacterial cytoplasm --- Certainty=0.4088(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 236/332 (71%), Positives = 276/332 (83%)

Query: 1 MKVLLYLEAE EYLKKSGIGRAIKHQEKALQIAGIDYTTNPTDDFDLVHMNTYGIRSWLLM 60  
 10 Sbjct: 1 MKVLLYLEAE YL+KSGIGRAIKHQ KAL + G +TTNP + +DLVH+NTYG++SWLLM 60

Query: 61 SKAKKTGKKVIMHGHSTEEDEFNRNFI SNL+SP FK YLC FY KADAIITPT YSK LI 120  
 15 Sbjct: 61 IKAQKAGKKVIMHGHSTEEDEFNRNFI SNLLSPWFKYLCHFYNKADAIITPTLYSKSLI 120

Query: 121 KAYGIKKPIFVLSNGIDLSRYQRSEKKESAFRHYFHL SKDDKVMGAGLYFMRKGIDQFV 180  
 20 Sbjct: 121 ESYGVKSPIFAVSNIDLEQYGADPKKEAFAFRYFDIKEGEKVMGAGLFFLRKGIDDFV 180

Query: 181 EVAAKMPDIRFIWFGETNKKVPIPRKVRQIVTKQHPSNVTFAGYIKGDVYEGAMSADAFF 240  
 25 Sbjct: 181 KVAQAMPDVRFIWFGETNKKVPIPAQVRQMVNGNHPKNLIFPGYIKGDVYEGAMTGADAFF 240

Query: 241 FPSREETEGIVVLEALASHQHVLRDIPVYHGWVTEDSVELATDVGFEKLDKVLGSGKS 300  
 25 Sbjct: 241 FPSREETEGIVVLEALASRQHLVLRDIPVYVGWVDQSSAELATDIPGFIEALKKVFSGAS 300

Query: 301 DKIEGYHVAESRSIERIAHELASVYQKVMEL 332  
 30 Sbjct: 301 NKVEAGYKVAQSRRLTVGHALVDVYKKVMEL 332

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 374

35 A DNA sequence (GBSx0405) was identified in *S.agalactiae* <SEQ ID 1221> which encodes the amino acid sequence <SEQ ID 1222>. Analysis of this protein sequence reveals the following:

Possible site: 16  
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.5487(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC35010 GB:AF055987 intracellular a-amylase [Streptococcus mutans]  
 Identities = 308/483 (63%), Positives = 378/483 (77%)

50 Query: 1 MTNELIMQAFEWYLPDGNHKKLEESISDLKKLGISKIWLPPAFKGTSSDDVG YGVYDL 60  
 MTNE +MQ FEWYLP+DG HW+ L E S LK +GISK+W+PPAFKGT S+DVGYGVYDL  
 Sbjct: 1 MTNETMMQYFEWYLPNDGKHWQHLAEDASHLKNIGISKVWMPAPFKGTGSNDVG YGVYDL 60

Query: 61 FDLGEFDQNGTIRTKYGRKEEYLKLIKSLKANGIKPFADIVLNHKANGDHKEKFQVIKVN 120  
 55 Sbjct: 61 YDLGEFNQNGTVRRTKYGSREDYLNALKEQEIMPISDIVLNHKANGDAKERFQVVKVN 120

Query: 121 PENRQEALSEPYEIEGWTGDFPGRQGEYNDFKWHWHFTGLDYDAKNNETDIFMIVGDN 180  
 P NRQE +SEPYEIEGWT F+FPGRQ Y+DFKWHWHFTG+DYDA +NE I+MI+GDN  
 60 Sbjct: 121 PSNRQEKISEPYEIEGWTQFNFPGRQDNYSDFKWHWHFTGVVDYDALHNENGIYIMIGDN 180

Query: 181 KGWADDDLIDDENGNFYDLYMYNDIDFKHPEVIKLNQDWAKWFIETTGIEGFRLDVAVKHID 240

-476-

KGWA + ID ENGN+DYLMY+DIDFKHPEV ++L+DW WF+ET+G+ GFRLLDA+KHID  
 Sbjct: 181 KGWASQENIDQENGNYDYLMYDDIDFKHPEVQEHLRDWVAVFLETSGVGGFRLDAIKHID 240

5 Query: 241 SYFIQTFFINDIRTKIKPDLEVFGEYWKSDQTSMDYLEATQFQFSLVDVTLHMNFDDASH 300  
 F+ FI IR +K DL VFGEYWK + DYL + QF L+DV LHM+ F+A  
 Sbjct: 241 KTFMAQFIRYIREHLKADLYVFGEYWKDSHFDITDYLHSDVLQFDLIDVMLHMSLFEEAGQ 300

10 Query: 301 QNRDFDMRTIFDDSLVIDNPEYAVTFVENHDTQSGQALESRVEDWFKPLAYGLILLRQQG 360  
 + DFD+ TI DDSL+ +P++AVTFV+NHD+Q GQALES V +WFKPLAYGLILLRQ+G  
 Sbjct: 301 KGSDFDLSTILDDSLMKSHDPFAVTFVDNHDQSGQALESTVAEWFKPLAYGLILLRQEG 360

15 Query: 361 TPCLFYGDYYGIQGEFGQPSFKEVIDKMAELRQNYVFGKQVDYFTHSNCIGWTCLGDEEH 420  
 PC+FYGDYYGI GEF Q SF+ V+DK+ +RQ +V+G + T NCIGWTCLGDEEH  
 Sbjct: 361 IPCVIFYGDYYGISGEFAQESFQTVLDKLLYIRQYHVYGSKKIILTMPNCIGWTCLGDEEH 420

Query: 421 NSCLAVVLTNGDQGWKHMEVGEIYAGKTFVDYLGNCQEVEVIGDDGWGDFLVESASISAW 480  
 +AV+++NG+ K M +GE K FVDYL NC +EV++ D GWGDF V+ AS+SAW  
 Sbjct: 421 PDGVAVIISNGEANCCKRMNMGEFNRNKVFVDYLNNTCEEVILDDQGWGDFPVQEASLSAW 480

20 Query: 481 VPK 483  
 V K  
 Sbjct: 481 VNK 483

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1223> which encodes the amino acid  
 25 sequence <SEQ ID 1224>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have a cleavable N-term signal seq.

30 ----- Final Results -----  
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the databases:

>GP:AAB00845 GB:M57692 alpha-cyclodextrin glycosyltransferase  
 [Thermoanaerobacterium thermosulfurigenes]  
 Identities = 356/710 (50%), Positives = 468/710 (65%), Gaps = 16/710 (2%)

40 Query: 7 KTYKLLTKSAVLLGLISFPLT--VSAADNASVTNKADFSTDTIYQIVTDRFNDGNTSNN 64  
 KT+KL+ + L L+ F LT + AA + +V+N ++STD IYQIVTDRF DGNTSNN  
 Sbjct: 3 KTFKLILVLMLSLTLV-FGLTAPIQAASDTAVSNVNYSTVDIYQIVTDRFVDGNTSNNP 61

45 Query: 65 KTDVFDKN--DLKKYHGGDWQGGIIAKIKDGYLTDMGISAIWISSPVENIDSIDPSN---G 119  
 D++D LKKY GGDWQGGII KI DGYLT MG++AIWIS PVENI ++ P + G  
 Sbjct: 62 TGDLYDPTHTSLKKYFGGDWQGGIINKINDGYLTGMGTAVIWIWISQPVENIYAVLPDSTFGG 121

50 Query: 120 SAAYHGYWAKDFFKTNQHFGEADFQQLVKVAHQHHIKVVIDFAPNHTSTAEKEGTTTFKE 179  
 S +YHGYWA+DF +TN +FG+ DFQ L+ AH H+IKV+IDFAPNHTS A + T+ E  
 Sbjct: 122 STSYHGYWARDFKRTNPYFGSFTDFQNLINTAHANNIKVIIDFAPNHTSPASETDPTYAE 181

55 Query: 180 DGALYKNGKLVGKFSDDKDKIFNHESWTFSTYENSIYHSMYGLADLNNINPKVDQYMKE 239  
 +G LY NG L+G +++D + F+H TDFS+YE+ IY +++ LADLN N +D Y+K  
 Sbjct: 182 NGRLYDNGTLLGGYTNDTNGYFHHYGGTDFSSYEDGIYRNLFDLADLNQONSTIDSYLKS 241

Query: 240 AIDKWLDLGVDGIRVDVAVKHMSQGWQKNWLSHIYEKHNVFVFGGEWFSGHTDDDDYDMITFA 299  
 AI WLD+G+DGIR+DAVKHM GWQKN++ I VF FGEWF G + D + T FA  
 Sbjct: 242 AIKVWLDMGIDGIRLDVAVKHMPFGWQKNFMDSILSYRVPVFTFGEWFLGTNEIDVNNTYFA 301

60 Query: 300 NNSGMGLLDLDFRANAIQLYTGSTFTMRDFYKVLNDRDQVTNEVTDQVTFIDNHDMERF 359  
 N SGM LLDLDFR+ +RQ++ +T TM ++++ N + D VTFIDNHD+RF  
 Sbjct: 302 NESGMSLLDFRFSQKVRQVFRD-NTDTMYGLDSMIQSTASDYNFINDMVTTFIDNHDMDRF 360

65 Query: 360 ATKVANNQTAVNQAYALLLTSRGVPIIYGYTEQYATGDKDPNNRGDMPSFNKESQAYKVI 419  
 + V QA A LTRGVP IYGYTEQY TG+ DP NR M SFN + AY VI

-477-

Sbjct: 361 YN--GGSTRPVEQALAFITLTSRGVPAIYYGTEQYMTGNGDPYNRAMMTSFNTSTTAYNVI 418

Query: 420 SKLAPLRKQNALAYGTTEQRWISDHVLVFERKFGNHVALVAINRDQINGYITITNAKTAL 479  
KLAPLRK N A+AYGTT+QRWI++ V ++ERKFGN+VALVAINR+ + Y IT TAL

5 Sbjct: 419 KKLAPLRKSNPAIAYGTTQQRWINNDVYIYERKFGNNVALVAINRNLSTSYNITGLYTAL 478

Query: 480 PQNSYKDKLEGLLGGQELIVGADGTISSFELGAGQVAVWVTEGEDKTPQLGDVDASVGIA 539  
P +Y D L GLL G + V +DG+++ F L AG+VAVW Y +P +G V ++ A

10 Sbjct: 479 PAGTYTDVLGGLNGNSISVASDGSVTFPTLSAGEVAVWQYVSSNSPLIGHVGPMTKA 538

Query: 540 GNKITISGQGFNGSKQVTFGEISAEILSWSDTLITLKVPTVPANYYNISVTTADKQTSN 599  
G ITI G+GFG + GQV FG + I+SW DT + +KVP+V YNIS+ T+ TSN

15 Sbjct: 539 GQTITIDGRGFGTTSGQVLFGSTAGTIVSWDDTEVKVKVPSVTPGKYNISLKTSSGATSN 598

Query: 600 SYQAFEVLTDKQIPVRLINDFKTVPGEQLYLMGDFEMGANDAKNAVGLPNNQTIAK 659  
+Y +LT QI VR ++N+ TV GE +YL G+V E+G D A+GP+FN Q + +

20 Sbjct: 599 TYNNINILTGNQICVRFVNNASTVYGENVYLTGNVAELGNWDTSKAIGPMFN--QVVYQ 656

Query: 660 YPNWFFDTHLPINKEIAVKLVKKDSIGNVLWT--SPETYSIKTGHEAQTI 707  
YP W++D +P I K +KK+ + W S TY++ + I

Sbjct: 657 YPTWYDVSVPA GTTIQFKFIKNG-NTITWEGGSNHTYTPSSSTGTVI 705

An alignment of the GAS and GBS proteins is shown below:

Identities = 112/509 (22%), Positives = 193/509 (37%), Gaps = 103/509 (20%)

25 Query: 18 GNHWKKLEESISD--LKKLGISKIWLPPAFKGTSSDDV-----GYGVYDLFDLGEFD 67  
G W+ + I D L +GIS IW+ + S D GY D F +

Sbjct: 79 GGDWQGIIAKIKDGYLTDMGISAIWISSPVENIDSIDPSNGSAAYHGYWAKDFFKTNQH- 137

30 Query: 68 QNGTIRTKYGRKEEYLKLIKSLKANGIKPFADIVLNHKANGDHKEKFQVIKVNPNRQEA 127  
+G + ++ +L+K + IK D NH + + +

Sbjct: 138 -----FGTEADFQQLVKVAHQHHIKVVIDFAPNHTSTAEKE----- 173

35 Query: 128 LSEPYEIEGWTGFDFFGRQGEYNDFKWHWHYHFTGLDYDAKNNETDIFMIVGDNKGWADDD 187  
G F Y + K G D K+ + +++ W D

Sbjct: 174 -----GTTFKEDGALYKNGK-----LVGKFSDDKDK-----IFNHESWTDFS 210

40 Query: 188 LIDDE--NGNFDYLMYNDIDFKHPEVIKNLQDWAKWFIETTGIEGFRLDVAVKHIDSYFIQ 245  
++ + + N+I+ K + +K D KW G++G R+DAVKH+ + +

45 Sbjct: 211 TYENSIYHSMYGLADLNNINPKVDQYMKEAID--KWL--DLGVDGIRVDVAVKHMSQGWQK 266

Query: 246 TFINDIRTKIKPDLEVFGEYWKSDQTSMDKYLEATQFQFSLVDVTLHMNFEDASHQ-NRD 304  
+++ I K ++ VFGE W S T D + T F + L F +A Q

Sbjct: 267 NWLSHIYE--KHNVFVFGE-WFSGHTD--DDYDMTTFANNSGMLLDFRANAIQLYTG 321

50 Query: 305 FDMRTIFDDSLVIDNPEYA-----VTFVENHDTQSGQALERSVEDWFKPLAYGLILLR 357  
F T+ D V++N + VTF++NHD + + + AY L LL

Sbjct: 322 FSTFTMRDFYKVLNDRDQVTNEVTDQVTFIDNHDMERFATKVANNQTAVNQ-AYAL-LLT 379

55 Query: 358 QQGTPLCFYGDYYGIQGE-----FGQPSFK-----EVIDKMAELR---QNYVFGKQVD 402  
+G P ++YG G+ PSF +VI K+A LR Q +G

Sbjct: 380 SRGVPNIIYGTQYATGDKDPNNRGMPSFNKESQAYKVISKLAPLRKQNALAYGTTEQ 439

Query: 403 YFTHSNICIGWTCLEDEHNSCLAVVLTNGDQ--GWKHMEVGEIYAGKTFVDYLGNC--EQ 458  
+ + + + +A+V N DQ G+ ++ D L Q

60 Sbjct: 440 RWISDHVL---VFERKFGNHVALVAINRDQINGYITITNAKTALPQNSYKDKLEGLLGGQ 495

Query: 459 EVVIGDDGW-GDFLVESASISAWVPKIEE 486  
E+++G DG F + + ++ W + E+

Sbjct: 496 ELIVGADGTISSFELGAGQVAVWVTEGED 524

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 375**

A DNA sequence (GBSx0406) was identified in *S.agalactiae* <SEQ ID 1225> which encodes the amino acid sequence <SEQ ID 1226>. This protein is predicted to be catabolite control protein A. Analysis of this protein sequence reveals the following:

```

5   Possible site: 29
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.2154 (Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9707> which encodes amino acid sequence <SEQ ID 9708> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:BAA88121 GB:AB028599 catabolite control protein A [Streptococcus
    bovis] (ver 3)
    Identities = 304/332 (91%), Positives = 320/332 (95%)

20 Query: 1  MNTDDTITIIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNVAVARGLA 60
    Sbjct: 1  MNTDDTITIIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNVAVARGLA 60

    Query: 61  SKKTTTVGVVIPNIANSYFSILARGIDIAAMYKYNIVLASSDEDDDDKEVNVVNTLFAKQ 120
25    Sbjct: 61  SKKTTTVGVVIPNIANSYFSILA+GIDIAAMYKYNIVLASSDEDDDDKEVNVVNTLFAKQ
    Sbjct: 61  SKKTTTVGVVIPNIANSYFSILAKGIDIAAMYKYNIVLASSDEDDDDKEVNVVNTLFAKQ 120

    Query: 121 VDGIIFMGHHLTEKIRAEFSRSRTPIVLAGTVDLHQPSVNIIDYKAAAVDVIDILAGNH 180
30    Sbjct: 121 VDGIIFMGHHLTEKIRAEFSRSRTPVVLACTVDLEHQLPSVNIIDYKAA DV+DILA N+
    Sbjct: 121 VDGIIFMGHHLTEKIRAEFSRSRTPVVLACTVDLEHQLPSVNIIDYKAAVADVVDILAKNN 180

    Query: 181 KDIAFVSGPLIDDINGKVRLAGYKEGLKKNLNFKEGLVFEANYRYAEGFALAQRVINAG 240
35    Sbjct: 181 KDIAFVSGPLIDDINGKVRLAGYKEGL+KN L+FKEGLVFEANY Y +G+ LAQRV+N+G
    Sbjct: 181 KDIAFVSGPLIDDINGKVRLAGYKEGLEKNNLSFKKEGLVFEANYNYKDGVELAQRVMNSG 240

    Query: 241 ATAAYVAEDELAAAGLLNGLFEAGKRPVDFEIIITSNDSPIAQYTRPNLTSISQPVYDLGA 300
40    Sbjct: 241 ATAAYVAEDELAAAGLLNGLF AGK+VPEDFEI+TSNDSPI YTRPNL+SISQPVYDLGA
    Sbjct: 241 ATAAYVAEDELAAAGLLNGLFAAGKVPEDFEIITSNDSPITSYTRPNLSSISQPVYDLGA 300

    Query: 301 VSMRMLTKIMHKEELEKEEKEIVLNHGIVKRGTT 332
45    Sbjct: 301 VSMRMLTKIM+KEELEKEEKEI+LNHG+ RGT
    Sbjct: 301 VSMRMLTKIMHKEELEKEEKEIILNHGILKRGTT 332

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1227> which encodes the amino acid sequence <SEQ ID 1228>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
50          bacterial cytoplasm --- Certainty=0.2154 (Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

55   Identities = 307/332 (92%), Positives = 320/332 (95%)

    Query: 1  MNTDDTITIIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNVAVARGLA 60
    Sbjct: 1  MNTDD +TIIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNVAVARGLA
    Sbjct: 1  MNTDDPLTIIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNVAVARGLA 60

```



-479-

Query: 61 SKKTTTGVGVIPNIANSYFSILARGIDDIAAMYKYNIVLASSDEDDDDKEVNVVNTLFAKQ 120  
 SKKTTTGVGVIPNIANSYFSILA+GIDDIAAMYKYNIVLASSDEDDDDKEVNVVNTLFAKQ  
 Sbjct: 61 SKKTTTGVGVIPNIANSYFSILAKGIDDIAAMYKYNIVLASSDEDDDDKEVNVVNTLFAKQ 120

5

Query: 121 VDGIIIFMGHHLTEKIRAEFSRSRTPIVLAGTVDLHQLPSVNIDYKAAAVDVIDILAGNH 180  
 VDGIIIFMGHHLTEKIRAEFSRSRTP+VLAGTVDL+HQLPSVNIDY+AA +V+DILA NH  
 Sbjct: 121 VDGIIIFMGHHLTEKIRAEFSRSRTPVVLAVLAGTVDLHQLPSVNIDYRAAVSNVVDILAENH 180

10

Query: 181 KDIAFVSGPLIDDINGKVRLAGYKEGLKKNLNFKEGLVFEANYRYAEGFALAQRVINAG 240  
 K IAFVSGPLIDDINGKVRLAGYKEGLK N L+FKEGLVFEANY Y EGF LAQRVIN+G  
 Sbjct: 181 KCIAFVSGPLIDDINGKVRLAGYKEGLKHNKLDKFEGLVFEANYSYKEGFELAQRVINS 240

15

Query: 241 ATAAYVAEDELAAGLLNGLFEAGKRVPEDFEIIITSNDSP+ QYTRPNL+SISQPVYDLGA 300  
 ATAAYVAEDELAAGLLNGLFEAGKRVPEDFEIIITSNDSP+ QYTRPNL+SISQPVYDLGA  
 Sbjct: 241 ATAAYVAEDELAAGLLNGLFEAGKRVPEDFEIIITSNDSPVQYTRPNLSSISQPVYDLGA 300

20

Query: 301 VSMRMLTKIMHKEELEKEIIVLNHGIVKRGTT 332  
 VSMRMLTKIM+KEELEKEI+LNHGI KRGTT  
 Sbjct: 301 VSMRMLTKIMHKEELEKEIILLNHGIKRGTT 332

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 376

25 A DNA sequence (GBSx0407) was identified in *S.agalactiae* <SEQ ID 1229> which encodes the amino acid sequence <SEQ ID 1230>. This protein is predicted to be PepQ (pepQ-2). Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1118(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC46293 GB:AF014460 PepQ [Streptococcus mutans]  
 Identities = 257/359 (71%), Positives = 304/359 (84%)

40 Query: 1 MSKLNRIHHLSVQAEALAVFSDPVTVNYLTGFFCDPHERQMFLFVYEDRDPILFVPALE 60  
 MSKL +I L E AV SDPV++NYLTGF+ DHER MFLF++ D++ +LF+P L+  
 Sbjct: 1 MSKLAQVQKLKQGI EA AVLSDPVSYNYLTGFYSDPHERLMFLFLFADQETLLFLPELD 60

45 Query: 61 VSRAKQSVFPFVFGYIDSENPWQKIASNLPSFSVSKVLA EFDNLNVTKFQGLQTVFDGHF 120  
 RAK + V GY+D ENP +KI + LP + SK+ EFDNLNVTKF+GL+T+F G F  
 Sbjct: 61 ALRAKSILDISVTGYLDFENPLEKIKTLLPKTNYSKIALEFDNLNVTKFKGLETIFSGQF 120

50 Query: 121 ENLTPYIQNMRLIKSRDEIEKMLVAGEFADKAVQVGFDNISLNNTE TDIIAQIEFEMKKQ 180  
 NLTP I MRLIKS DEI+K+L+AGE ADKAVQ+GFD+ISLN TETDIIAQIEFEMKK  
 Sbjct: 121 TNLTPLINRMRLIKSADEIQKLLIAGELADKAVQIGFDSISLNATETDIIAQIEFEMKKL 180

55 Query: 181 GINKMSFDMVLTGNNAANPHGIPGTNKIENNALLLFDLGVETLGYTSDMTRTVAVGKPD 240  
 G++KMSF+TMVLTG+NAANPHG+P ++KIENN LLLFDLGV E+ GY SDMTRTVAVG+PD  
 Sbjct: 181 GVDKMSFETMVL TGSNAANPHGLPASHKIENNHL LLLFDLGVESTGYVSDMTRTVAVGQPD 240

60 Query: 241 QFKKDIYHLCLEAHQA AIDFIKPGVLASEVDAAARNVIEKAGYGQYFNHRLGHGLGMDVH 300  
 QFKKDIY++CLEA A+DFIKPGV A++VDAAAR+VIEKAGYG YFNHRLGHG+GM +H  
 Sbjct: 241 QFKKDIYNICLEAQLTALDFIKPGVSAAQVDAARSVIEKAGYGDYFNHRLGHGIGMGLH 300

Query: 301 EFPSIMAGNDMEIQEGMCFVSVEPGIYIPDKVGVRIEDCGYVTKTGFEVFTKTPKELLYF 359  
 EFPSIMAGNDM ++EGMCFVSVEPGIYIP+KVGVRIEDCG+VTK GFEVFT+TPKELLYF

-480-

Sbjct: 301 EFPSIMAGNDMLLEEGMCFSEVPGIYIPEKVGVRIEDCGHVTKNGFEVFTQTPKELLYF 359

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1231> which encodes the amino acid sequence <SEQ ID 1232>. Analysis of this protein sequence reveals the following:

5           Possible site: 58  
 >>> Seems to have no N-terminal signal sequence  
       INTEGRAL      Likelihood = -0.90      Transmembrane    42 - 58 ( 42 - 59)  
 10       ----- Final Results -----  
           bacterial membrane --- Certainty=0.1362(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

15       >GP:AAC46293 GB:AF014460 PepQ [Streptococcus mutans]  
       Identities = 264/359 (73%), Positives = 304/359 (84%)  
 Query: 1    MTKLDQIRLYLDQKGAEIAIFSDPVTINYLTFGFFCDPHERQLFLFVYHDLAPVLFVPALE 60  
           M+KL QI   L ++G E A+ SDPV+INYLTFG+ D PHER +FLF++ D   +LF+P L+  
 20   Sbjct: 1   MSKLAQIVQKLKKQIEAAVSDPVSINYLTFGFSDPHERLMFLFLFADQETLLFLPELD 60  
 Query: 61   VARASQAISFPVFGYVDSNPWEKIKAVLPNTAAKTIYAEPDHLNVNKFHGLQTIFSGQF 120  
           RA   +   V GY+D ENP EKIK +LP T   I   EFD+LNV KF GL+TIFSGQF  
 25   Sbjct: 61   ALRAKSILDISVTGYLDNFENPLEKIKTLLPKTNYSKIALEFDNLNVTKFGLETIFSGQF 120  
 Query: 121   NNLTPYVQGMRLVKSADIEINKMMIAGQFADKAVQVGFDNISLDTATETDVIAQIEFEMKKQ 180  
           NLTP +   MRL+KSADEI K++IAG+ ADKAVQ+GFD+ISL+ATETD+IAQIEFEMKK  
 30   Sbjct: 121   TNLTPLINRMRLIKSADEIQKLLIAGELADKAVQIGFDSISLNATETDIIAQIEFEMKKL 180  
 Query: 181   GIHKMSFDTMVLGTGNNANPHGIPGTNNIENNALLLFDLGVETLGYTSDMTRTVAVGQPD 240  
           G+ KMSF+TMVLGT+NAANPHG+P ++ IENN LLLFDLGV+ GY SDMTRTVAVGQPD  
 35   Sbjct: 181   GVDKMSFETMVLGTGNNANPHGLPASHKIENNHLLLFDLGVESTGYVSDMTRTVAVGQPD 240  
 Query: 241   QFKIDIYNLCLEAQLAAIDFIKPGVTAAQVDAAARQVIEKAGYGEYFNHRLGHGIGMDVH 300  
           QFK DIYN+CLEAQL A+DFIKPGV+AAQVDAAAR VIEKAGYG+YFNHRLGHGIGM +H  
 40   Sbjct: 241   QFKKDIYNLCLEAQLTALDFIKPGVSAAQVDAAARSVIEKAGYGDYFNHRLGHGIGMGLH 300  
 Query: 301   EFPSIMAGNDLVLEEGMCFSEVPGIYIPEKVGVRIEDCGHVTKNGFEVFTHTPKELLYF 359  
           EFPSIMAGND++LEEGMCFSEVPGIYIP KVGVRIEDCGHVTKNGFEVFT TPKELLYF  
 40   Sbjct: 301   EFPSIMAGNDMLLEEGMCFSEVPGIYIPEKVGVRIEDCGHVTKNGFEVFTQTPKELLYF 359

An alignment of the GAS and GBS proteins is shown below:

Identities = 288/361 (79%), Positives = 325/361 (89%)  
 45   Query: 1   MSKLNLRIRHHLHSVQAEIAVSDPVTVNYLTGFFCDPHERQMFLFVYEDRDPILFVPALE 60  
           M+KL++IR +L   AELA+FSDPVT+NYLTGFFCDPHERQ+FLFVY D   P+LFPVPALE  
   Sbjct: 1   MTKLDQIRLYLDQKGAEIAIFSDPVTINYLTFGFFCDPHERQLFLFVYHDLAPVLFVPALE 60  
 Query: 61   VSRAKQSVFPVFGYIDSENWPQKIASNLPSFSVSKVLAEPDNLNVTKFQGLQTVFDGHF 120  
           V+RA Q++ FFPVFGY+DSENWP+KI + LP+ +   + AEPD+LNV KF GLQT+F G F  
 50   Sbjct: 61   VARASQAISFPVFGYVDSNPWEKIKAVLPNTAAKTIYAEPDHLNVNKFHGLQTIFSGQF 120  
 Query: 121   ENLTPYIQNMRLIKSRDEIEKMLVAGEFADKAVQVGFDNISLNTTETDIIAQIEFEMKKQ 180  
           NLTPY+Q MRL+KS DEI KM++AG+FADKAVQVGFDNISL+ TETD+IAQIEFEMKKQ  
 55   Sbjct: 121   NNLTPYVQGMRLVKSADIEINKMMIAGQFADKAVQVGFDNISLDTATETDVIAQIEFEMKKQ 180  
 Query: 181   GINKMSFDTMVLGTGNNANPHGIPGTNKIENNALLLFDLGVETLGYTSDMTRTVAVGKPD 240  
           GI+KMSFDTMVLGTGNNANPHGIPGTN IENNALLLFDLGVETLGYTSDMTRTVAVG+PD  
 60   Sbjct: 181   GIHKMSFDTMVLGTGNNANPHGIPGTNNIENNALLLFDLGVETLGYTSDMTRTVAVGQPD 240  
 Query: 241   QFKKDIYHLCLEAHQAIDFIKPGVLASEVDAAARNVIEKAGYGYFNHRLGHGLGMDVH 300  
           QFK DIY+LCLEA AAIDFIKPGV A++VDAAAR VIEKAGYG+YFNHRLGHG+GMDVH  
   Sbjct: 241   QFKIDIYNLCLEAQLAAIDFIKPGVTAAQVDAAARQVIEKAGYGEYFNHRLGHGIGMDVH 300

-481-

Query: 301 EFPSIMAGNDMEIQEGMCFSEPGIYIPDKVGVRIEDCGYVTKTGFEVFTKTPKELLYFEG 361  
 EFPSIMAGND+ ++EGMCFSEPGIYIP KVGVRIEDCG+VTK GFEVFT TPKELLYFEG  
 Sbjct: 301 EFPSIMAGNDLVLEEGMCFSEPGIYIPGKVGVRIEDCGHVTKNGFEVFTHTPKELLYFEG 361

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 377

A DNA sequence (GBSx0408) was identified in *S.agalactiae* <SEQ ID 1233> which encodes the amino acid sequence <SEQ ID 1234>. Analysis of this protein sequence reveals the following:

10 Possible site: 14  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 15 bacterial cytoplasm --- Certainty=0.3629(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 378

A DNA sequence (GBSx0409) was identified in *S.agalactiae* <SEQ ID 1235> which encodes the amino acid sequence <SEQ ID 1236>. This protein is predicted to be beta-hexosamidase A precursor. Analysis of this protein sequence reveals the following:

Possible site: 47  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 30 bacterial cytoplasm --- Certainty=0.3279(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

35 >GP:CAB11942 GB:Z99104 alternate gene name: yzbA-similar to  
 beta-hexosaminidase [Bacillus subtilis]  
 Identities = 151/602 (25%), Positives = 268/602 (44%), Gaps = 69/602 (11%)

Query: 26 INEMTLDEKIGQLF-----FNMGASRSEELYTDVLDRIYHIAAVRYNRGSSSEIYDQNL- 78  
 40 +N M+LDEK+GQ+ + S + LT + D +Y G ++ +N+  
 Sbjct: 39 VNRMSLDEKLGQMLMPDFRNWQKEGESSPQALTKMNDENVASLVKKYQFGGII-LFAENVK 97

Query: 79 -----ILQTKSKLPMLIAANTEAGGDGAVTDGTVKGDEIKVAATNDPKYAYEMG 127  
 + K+P++++ + E G + +GT + + A AY+ G  
 45 Sbjct: 98 TTKQTVQLTDDYQKASPKIPLMLSIDQEGGIVTRLGEGTNFFGNMALGAARSINAYQTG 157

Query: 128 RIAGMEASAVGCNASFSPIVDLTRNWRNPITASRNWGANVDQIISLSKEYMKGIMQYNIV 187  
 I G E SA+G N FSP+VD+ N NP+I R++ +N + L MKG+ + +I  
 Sbjct: 158 SIIGKELSALGINTDFSPVVDINNNPDNPVIGVRSFSSNRELTSRLGLYTMKGLQRQDIA 217

50 Query: 188 PFAKHFPDGDIDERDHLSFASNPMKBEWMSTFGRIYGELADAGLPGVMAGHIHLNVE 247  
 KHFP G + D H +E + + DAG VM H+ P +  
 Sbjct: 218 SALKHFPFGHGD TDVDSHYGLPLVSHGQERLREVELYPFQKAIDAGADMVMTAHVQFPAPD 277

-482-

Query: 248 KEMHPER--DLDDMLPASLNKTLLEDELLRGELGYNGAIVTDASHMVGMTASMARRDLLPT 305  
 + + D ++PA+L+K ++ LLR E+G+NG IVTDA +M + + + +  
 Sbjct: 278 DTTYKSKLDGSDILVPATLSKKVMTGLLRQEMGFNGVIVTDALNMKAIADHFGQEEAVVM 337

5 Query: 306 AIEAGCDLFLF---FNDDPED-----IQWMKEGYEKGILTEERLHDALRRTLGLKAKLG 356  
 A++AG D+ L E+ IQ +KE + G + E+++++++ R + LK K G  
 Sbjct: 338 AVKAGVDIALMPASVTSLKKEQKFARVIQALKEAVKNGDIPEQQINNSVERIISLKIKRG 397

10 Query: 357 LHNVEGRRQELFMPK-DKAMALINTLESQKIADEVADKAVTLVKDKQKDIFPVNPERYRH 415  
 + Y R + K KA ++ + + K ++A+KAVT++K++Q + P P++  
 Sbjct: 398 M--YPARNSDSTKEKIAKAKKIVGSKQHLKAEKKLAEKAVTVLKNEQHTL-PFKPKKGSR 454

15 Query: 416 ILLVNVVEGYKGGFGAMIAGNKQRASDYMKE-----LLEARGHEVTVWSTEERIMKLPQ 469  
 IL+V + A +Q D +K L V+++ E+ +K  
 Sbjct: 455 ILIV-----APYEEQTASIEQTIHDLIKRKKIKPVSLSKMNFASQVFKTEHEKQVK--- 505

20 Query: 470 EERAAAIAINVYAQK-QPIANLTHEYDLIINLVDVNAGGTTQRIIWPAAGKTPDQPFYVHE 528  
 E I Y K P+ N D +I+ D + + ++P A + H  
 Sbjct: 506 -EADYIITGSYVVKNDPVVN-----DGVID--DTISDSSKWATVFPRA---VMKAALQHN 554

Query: 529 IPSIVISVQHAFALADMPQVGTINAYD-----GLPSTISAVVAKLAGESEFTGVSP 580  
 P +++S+++ + A+ + I Y L I A V + G+++ G P  
 Sbjct: 555 KPFVLMSLRNPYDAANFEEAKALIAVYGFKGYGANGRYLQPNIPAGVMAIFGQAKPKGTLP 614

25 Query: 581 VD 582  
 VD  
 Sbjct: 615 VD 616

No corresponding DNA sequence was identified in *S.pyogenes*.

30 A related GBS gene <SEQ ID 8565> and protein <SEQ ID 8566> were also identified. Analysis of this protein sequence reveals the following homology to a lipoprotein, with homology with the following sequences in the databases:

29.5/52.3% over 422aa

35 EGAD|20114| hypothetical 70.6 kd protein in feua 5'region precursor Insert characterized  
 SP|P40406|YBBD\_BACSU HYPOTHETICAL 70.6 KDA LIPOPROTEIN IN FEUA-SIGW INTERGENIC REGION  
 PRECURSOR (ORF1). Insert characterized  
 GP|1944006|dbj|BAA19499.1||AB002150 YbbD Insert characterized

40 GP|438455|gb|AAA64351.1||L19954 possible N-terminal signal sequence; mature protein may  
 be membrane-anchored and start at Cys-17. 17.5% identity  
 over 354-aa overlap with Candida pelliculosa beta-glucosidase.; putative Insert  
 characterized  
 GP|2632433|emb Insert characterized

45 ORF00431(367 - 1557 of 2388)  
 EGAD|20114|BS0166(36 - 458 of 642) hypothetical 70.6 kd protein in feua 5'region precursor  
 {Bacillus subtilis} SP|P40406|YBBD\_BACSU HYPOTHETICAL 70.6 KDA LIPOPROTEIN IN FEUA-SIGW  
 INTERGENIC REGION PRECURSOR (ORF1). GP|1944006|dbj|BAA19499.1||AB002150 YbbD {Bacillus  
 50 subtilis} GP|438455|gb|AAA64351.1||L19954 possible N-terminal signal sequence; mature  
 protein may be membrane-anchored and start at Cys-17. 17.5% identity over 354-aa overlap  
 with Candida pelliculosa beta-glucosidase.; putative {Bacillus subtilis} GP|2632433|emb  
 %Match = 9.6  
 %Identity = 29.5 %Similarity = 52.2  
 55 Matches = 119 Mismatches = 183 Conservative Sub.s = 92

114 144 174 204 234 264 294 324  
 LMVGDSLGLDLAAAEQNGIAFYFVLVGKEVKSWEILREDIGEAFAGQFEQQRQKESINTFWANLDN\*\*KG\*AMTHLVDLT

60 MRPVFPLILSAVLFLSCFFGA  
 10 20

354 384 414 426 456 486 528  
 KKPFNLNQEAIEWIEKTINEMTLDEKIGQLFF-----NMGASRSEYLTVDLDRYHIAAVRYNRGS-----SSEIYDQ



Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 380

A DNA sequence (GBSx0411) was identified in *S.agalactiae* <SEQ ID 1239> which encodes the amino acid sequence <SEQ ID 1240>. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2279(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9705> which encodes amino acid sequence <SEQ ID 9706> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC21726 GB:U32690 oxidoreductase [Haemophilus influenzae Rd]  
Identities = 197/271 (72%), Positives = 229/271 (83%)

Query: 26 NKVVVITGAGGVLCGYMAKEFAKAGAKVALLDLNQEAAQTFAD EIVEEGGI AKAYKANVL 85  
NK+++ITGAGGVLC ++AK+ A A +ALLDLN EAA A EI + GG AKAYK NVL  
Sbjct: 15 NKLIITGAGGVLC SFLAKQLAYTKANIALLDLNF EADKVAKEINQSGGKAKAYKTNVL 74

Query: 86 SKENLEEVHQAVLEDLGPTDILVNGAGGNNPKATTDNEFH ELDLPSETKTFFELDEAGIS 145  
EN++EV + D G DIL+NGAGGNNPKATTDNEFH+ DL T+TFF+LD++GI  
Sbjct: 75 ELENIEKVRNQIETDFGTC DILINGAGGNNPKATTDNEFHQFDLNETTRTFFDLKSGIE 134

Query: 146 FVFNLNLYLGTLPTQVFAQDMVGREGANIINISSMNAFTPLTKIPAYSGAKAAISNFTQW 205  
FVFNLNLYG+LLPTQVFA+DM+G++GANIINISSMNAFTPLTKIPAYSGAKAAISNFTQW  
Sbjct: 135 FVFNLNLYGSLPTQVFAKDLGKQGANIINISSMNAFTPLTKIPAYSGAKAAISNFTQW 194

Query: 206 LAVHFSKVGIRCNAIAPGFLVTNQNRSLLFTEDGQPTARAEKILNNTPMGRFGEASELIG 265  
LAV+FSKVGIRCNAIAPGFLV+NQN +LLF +G+PT RA KIL NTPMGRFGE+ EL+G  
Sbjct: 195 LAVYFSKVGIRCNAIAPGFLVSNQNLLALFDTEGKPTDRANKILTNTPMGRFGESEELLG 254

Query: 266 GLFFLADEKSSSFVNGVVLPIDGGFAAYSGV 296  
L FL DE S+FVNGVVLP+DGGF+AYSGV  
Sbjct: 255 ALLFLIDENYSAFVNGVVLPVDGGFSAYSGV 285

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1241> which encodes the amino acid sequence <SEQ ID 1242>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0358(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 77/279 (27%), Positives = 125/279 (44%), Gaps = 19/279 (6%)

Query: 18 MSKTITFTNKVVVITGAGGVLCGYMAKEFAKAGAKVALLDLNQEAAQTFAD EIVEEGGIA 77  
M + K+ +ITGA + +AK +A+AGA + D+ QE E G A  
Sbjct: 1 MENMFSLQGKIALITGASYGIGFEIAKAYAQA GATIVFNDIKQELVDKGLAAYRELGIEA 60

Query: 78 KAYKANVLSKENLEEVHQAVLEDLGPTDILVNGAGGNNPKATTDNEFH ELDLPSETKTFF 137

-485-

```

      Y +V + ++++ + +++G DILVN AG
Sbjct: 61 HGYVCDVTDEAGIQQMVSQIEDEVGAIDILVNNAG-----IIRRTPML 103

Query: 138 ELDEAGISFVFNLNYLGTLLPTQVFAQDMVGREGANIINISSMNAFTPLTKIPAYSGAKA 197
5      E+      V +++      + ++      M+ +      IINI SM +      + AY+ AK
Sbjct: 104 EMAAEDFRQVIDIDLNAPFIVSKAVLPSMIAGHGKIINICSMMSSELGRETVSAYAAAKG 163

Query: 198 AISNFTQWLAVHFSKVGIRCNAIAPGFLVTNQNRSLLFTE-DGQPTARAEKILNNTPMGR 256
      + T+ +A F + I+CN I PG++ T Q L + DG + I+ TP R
10 Sbjct: 164 GLKMLTKNIASEFGEANIQCNGIGPGYIATPQTAPLRERQADGSRHPFDQFIIAKTPAAR 223

Query: 257 FGEASELIGGLFFLADEKSSSFVNGVVLPIDGGFAAYSG 295
      +G +L G FLA + +S+FVNG +L +DGG AY G
15 Sbjct: 224 WGTTEDLAGPAVFLASD-ASNFNHGHILYVDGGILAYIG 261

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 381

A DNA sequence (GBSx0412) was identified in *S.agalactiae* <SEQ ID 1243> which encodes the amino acid sequence <SEQ ID 1244>. This protein is predicted to be D-mannonate dehydrolase (uxuA). Analysis of this protein sequence reveals the following:

```

Possible site: 42
>>> Seems to have no N-terminal signal sequence

```

```

25 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3188(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

30 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:BAB04425 GB:AP001509 D-mannonate dehydrolase [Bacillus halodurans]
Identities = 202/343 (58%), Positives = 261/343 (75%)

```

```

35 Query: 1 MEMSFRWYGEDDPVTLENIGQIPTMKGIVTAIYDVPVGEVWSRERIQQLEKEVEAAGLKI 60
      M ++ RW+G D V LE I QIP MKGIV+AIYDV VG VW +E+I LK +E GL +
Sbjct: 1 MRLTMRWFGPSDKVKLEYIKQIPGMKGIVSAIYDVAVGGVWPKEKILALKNNIERHGLTL 60

Query: 61 SVIESVPVHEDIKLRPTRDLLIDNYIQTVKNLAAEGIDTICYNFMPVFDWTRTDLAYQY 120
      VIESVPVHEDIKLG+PTRD I+NY QT+++LA GIDT+CYNFMPVFDWTR+ L ++
40 Sbjct: 61 DVIESVPVHEDIKLGKPTRDRYIENYKQTLRHLAECGIDTVCYNFMPVFDWTRSQLDFKL 120

Query: 121 PDGSTALIFDETIVSKMDPVNGELSLPGWDASYSKEEMKAIMDAYAEIDEEKLWENLTYF 180
      DGS ALI++E V + +P++GEL LPGWD SY E +K ++ AY +I EE LW++LTYF
45 Sbjct: 121 EDGSEALIYEEDVISRTNPLSGELELPGWDTSYENESLKGVLQAYKKISEEDLWDHLTYF 180

Query: 181 IKRIIPEAEAVGVKMAIHPDDPPYSIFGLPRIITGLEAIERFVKLYDSKNGITLCVGSY 240
      ++ I+P A+ VG+KMAIHPDDPP+SIFGLPRI+T +ER + LYDS ++GIT+C GS
Sbjct: 181 VQAIMPVADEVGKMAIHPDDPPWSIFGLPRIVTNKANLERLLSLYDSPNHGITMCSGSL 240

50 Query: 241 ASDPQNDVLEISRRAFELDRVNFVHARNIKLGDGKSFKESAHPSYGSIDMYEVIKLCHE 300
      ++ ND+ E+ R R++F HARNIK +SF+ESAH SE GS++M ++K H+
Sbjct: 241 GANEANDLPEMIRHFGGQRIHFAHARNIKRTGPRSFQESAHLSEAGSVNMVAMLKAYHD 300

55 Query: 301 FGFEGAIRPDHGRMIWGETGRPGYGLYDRALGATYVSGLYEAV 343
      GF G +RPDHGRMIWGE GRPGYGLYDRALGATY++G++EAV
Sbjct: 301 IGFTGPLRPDHGRMIWGEKGRPGYGLYDRALGATYLNIGIWEAV 343

```

No corresponding DNA sequence was identified in *S.pyogenes*.

-486-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 382**

A DNA sequence (GBSx0413) was identified in *S.agalactiae* <SEQ ID 1245> which encodes the amino acid sequence <SEQ ID 1246>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2447(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 383**

A DNA sequence (GBSx0414) was identified in *S.agalactiae* <SEQ ID 1247> which encodes the amino acid sequence <SEQ ID 1248>. This protein is predicted to be uronate isomerase. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3066(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04424 GB:AP001509 uronate isomerase [Bacillus halodurans]

Identities = 215/465 (46%), Positives = 294/465 (62%), Gaps = 7/465 (1%)

Query: 3 FNTETFMLKNQAAIQLYEE-VKRQPIFDYHCHLDPKDIFEDHIFDNIVDLWLGGDHYKWR 61

F +E F+L N+ +LY K PI DYHCHL P++I+E+ F+N+ WLGGDHYKWR

Sbjct: 4 FLSEDPLLNMNEYDRELYYTFKNMPCIDYHCHLSPQEIWENKPFENMTKAWLGGDHYKWR 63

Query: 62 LMRANGISEAETGPASNLEKFKAFARTLERAYGNPVYHWSAMELKNVFGVNEILTESNA 121

MR NG+ E ITG A + EK F A+A+T+ + GNP+YHW+ MELK F ++ L E+N

Sbjct: 64 AMRLNGVREEFITGGAPDKEKFLAWAKTVPKTIGNPLYHWHMELKTYFHFHQPLDETNG 123

Query: 122 EEIYHRLNHFLKEHKISPRRLIADSKVMFIGTDDHPLDTLEWHKKLADESFKTVVAPTF 181

E ++ N L++ +PR LI S V IGTDD P D+L +H+KL AD++F V PTF

Sbjct: 124 ENVWDACNRLQLQEAFTPRALIERSNVRAIGTTDDPTDSLHYHQLQADDTFHVKVIPTF 183

Query: 182 RPDEAF-IEHRHFVDFITKLGDITQKEITDFSTFIAAMEERIAFYAQNCGCRASDISFTEI 240

RPD A IE F D++ KL D+T + + F+ A++ER+ +F ++GCR+SD TE+

Sbjct: 184 RPDGALKIEQDSFADWVAKLSDVTGESLDTLDAFLHALKERLTFFDEHGCRSSDHDMTEV 243

Query: 241 VFEQTDELELNDLFNKVCEGYIPNQSEISKWQTAVFMELCRLYKKYGFVTVQVHFGALRNN 300

F + +E E +F K + E K++T + L + Y G+V Q H G +RNN

Sbjct: 244 PFVEVNEQEAQHIFRKRLANEGLTKVENEKYKTFMLTWLGKEYAARGWVMQWHIGVMRNN 303

Query: 301 HSTIFEKLGAADVGVDSLGD-QVALTVNMNRLDLSLVKKDSLPPKMIWYNLNPAYNIAVANT 359



-487-

+S + KLG D G DS+GD Q+A +LLD L K+ +LPK I Y +NP N +A+  
 Sbjct: 304 NSRMLHLKLPDGTGFDSIGDQIAHAT--AKLLDLLDKQALPKTILYCVNPNANYILASM 361

Query: 360 LANFQANELGVRSYLQFGAGWWFADTKLGMISQMNALAEQGMLANFIGMLTDSRSFLSYQ 419  
 + NF E GVR +QFG+ WWF D GM Q+ LA G+L+NFIGMLTDSRSFLSY  
 Sbjct: 362 IGNF--TESGVRGKVQFGSAWWFNDHIDGMRRQLTDLASVGLLSNFIGMLTDSRSFLSYP 419

Query: 420 RHDYFRILCTYLGEWIEEGEVPEDYQALGSMADIAQNAVNYF 464  
 RHDYFRILC +G WI+EG++P D + G + +DI Y N V+YF  
 Sbjct: 420 RHDYFRILCQLIGSWIKEGQLPPDMERWGQIVQDICYNVVDYF 464

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 15 Example 384

A DNA sequence (GBSx0415) was identified in *S.agalactiae* <SEQ ID 1249> which encodes the amino acid sequence <SEQ ID 1250>. This protein is predicted to be 2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate al. Analysis of this protein sequence reveals the following:

Possible site: 43  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3883(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9703> which encodes amino acid sequence <SEQ ID 9704> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD35160 GB:AE001693 2-dehydro-3-deoxyphosphogluconate  
 aldolase/4-hydroxy-2-oxoglutarate aldolase [Thermotoga maritima]  
 Identities = 93/199 (46%), Positives = 125/199 (62%), Gaps = 6/199 (3%)

Query: 37 KNNYFFAVIRGKSSDALEIAKHAILGGIRNIEVTFSTPEASKVIKQLSDDFKNNKEIIV 96  
 K + AV+R S E+A E A GG+ IE+TF+ P+A VIK+LS F K I+  
 Sbjct: 8 KKKKIVAVLRANSVEEAKKALAVFEGGVHLIEITFTVPDADTVIKELS--FLKEKGAI 65

Query: 97 GAGTVMTELAKEAIDAGAKFLVSPHFDSDIANLANENKVYYPGCGATATEIVVARKYKC 156  
 GAGTV + E ++A+++GA+F+VSPH D +I+ E V+Y PG T TE+V A K  
 Sbjct: 66 GAGTVTSVEQCRKAVESGAEFIVSPHLDDEEISQFCKEKGVFYMFGVMTPTLVKAMKLG 125

Query: 157 QIIKLFPGGVVGPFGFIKDIHGPIPDVDLMPSGGVSVSNVVEWRKAGAVAVGVGSALSSKV 216  
 I+KLFPFG VVGP F+K + GP P+V +P+GGV++ NV EW KAG +AVGVGSAL  
 Sbjct: 126 TILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGT 185

Query: 217 ATEGYDSVTAKIQAQFVSAL 235  
 D V + AK FV +  
 Sbjct: 186 P----DEVREKAKAFVEKI 200

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1251> which encodes the amino acid sequence <SEQ ID 1252>. Analysis of this protein sequence reveals the following:

Possible site: 29  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1039(Affirmative) < succ>

-488-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

5       Identities = 82/204 (40%), Positives = 132/204 (64%)

Query: 32   MLNLQKNNYFFAVIRGKSSDALEIAKHAILGGIRNIEVTFSTPEASKVIKQLSDDFKNN 91  
           +L +LK N    V+RG+SSE+AL + +I GGI+ IEVT++ P AS+VI QL++ FK +  
 10   Sbjct: 6   ILTKLKANRLVLVVRGESSEALACSLASIEGGIKTIEVTTYTNPFASEVIGQLAERFKED 65

Query: 92   KEIIVGAGTVMTTELAKAIDAGAKFLVSPHFDSDIANLANENKVYFPGCATATEIVVA 151  
           E+++GAGTV+    A++AI AGA+F+V P+F+ +A + +   + Y PGC T E+V A  
 15   Sbjct: 66   PEVLIGAGTVLDDVTARQAILAGAQFIVGPNFNRAVALICHRYSSIPYLPGCMTVNEVVTA 125

Query: 152   RKYKQCIILKLPFGVVGPGFIKDIHGPIPDVDLMPSGGVSVSNVVEWRKAGAVAVGVGSA 211  
           +    ++K+FPG VG FI+ I P+P V++M +GGVS N+ +W AG +G+G  
 20   Sbjct: 126   LESGVDMVKIFPGSTVGISFIRAISPLPQVEVMVTGGVSSDNLKDWLAAGVDVLGIGGE 185

Query: 212   LSSKVATEGYDSVTKIAKQFVSAL 235  
           +   + + Y+ +TK A ++ +L  
 25   Sbjct: 186   FNQLASQKQYNLITKKAHYIKSL 209

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 25   Example 385

A DNA sequence (GBSx0416) was identified in *S.agalactiae* <SEQ ID 1253> which encodes the amino acid sequence <SEQ ID 1254>. This protein is predicted to be pyruvate dehydrogenase complex repressor. Analysis of this protein sequence reveals the following:

30       Possible site: 26  
        >>> Seems to have no N-terminal signal sequence

----- Final Results -----

          bacterial cytoplasm --- Certainty=0.2827(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 35       bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12044 GB:Z99105 similar to transcriptional regulator (GntR family) [Bacillus subtilis]

40       Identities = 67/225 (29%), Positives = 119/225 (52%), Gaps = 17/225 (7%)

Query: 3    RPLVEQTADRLHLILEREYPVGAKLPNEYELAEDLDVGRSTIREAVRSLATRNILEVRQ 62  
           + L +Q +R++HL+    +   G KLP E EL + L V R +REA+ SL T ++ +  
 45   Sbjct: 16   KTLAKQVIERIVHLLSSGQLRAGDKLPTEMELMDILHVSRLPVLREALSSLETGLGVITRKT 75

Query: 63   GSGTYISSKKGVSIEDPLGFSLIKDTDRITSDLFELRLILLEPRIAELVAYRITDDQLQLLE 122  
           GTY + K G+   P    L   TD L + + E R+ LE + + A +I +++LQ L+  
 50   Sbjct: 76   RGGTYFNDKIGM--QPFSVMLALATDNLPA-IIEARMALELGLVTIAAEKINEEELQRLQ 132

Query: 123   KLVGDIEDAV--HAGDPKHLLEDVEFHSMLAKYSGNIAMDSLLPVINQSIHLINANYTNR 180  
           K + DI ++   H G+    D EFH ++A + N ++ ++   QS+ + +A ++  
 55   Sbjct: 133   KTIDDIANSTDNHYGE-----ADKEFHRIIALSANNPVVEGMI-----QSLITHAKIDSQ 183

Query: 181   ---QMKSDSLEAHREIIKAIREKNPVAAHDAMLHMTSVRRSALK 222  
           + +   ++E H++I A+ +++P AH M H+ VR LK  
 60   Sbjct: 184   IPYRERDVTVEYHKKIYDALAQRDPYKAHYHMYEHLKFVRDKILK 228

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1255> which encodes the amino acid sequence <SEQ ID 1256>. Analysis of this protein sequence reveals the following:

-489-

Possible site: 54

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2161(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 24/51 (47%), Positives = 35/51 (68%)

Query: 22 YPVGAKLPNEYELAEDLDVGRSTIREAVRSLATRNILEVRQSGTYISSKK 72  
 +P+G++LP+E LAE V R T+R+A+ L ILE R GSGTY++S +  
 15 Sbjct: 30 WPIGSRPLPSEHRLAEHFTVSRMTLRQAITLLVEEGILERRIGSGTYVASHR 80

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 386**

A DNA sequence (GBSx0417) was identified in *S.agalactiae* <SEQ ID 1257> which encodes the amino acid sequence <SEQ ID 1258>. Analysis of this protein sequence reveals the following:

Possible site: 31

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2178(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9701> which encodes amino acid sequence <SEQ ID 9702> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA58911 GB:X84105 gluceronidase [synthetic construct]  
 Identities = 258/602 (42%), Positives = 357/602 (58%), Gaps = 31/602 (5%)

35 Query: 23 MLYPLLTKTRNTYDLGGIWNFKLGEHNPN-----ELLPSDEVMVIPTSFNDLMVSKEK 75  
 ML P+ T TR L G+W F L N L + +P SFND +  
 Sbjct: 1 MLRPVETPTREIKKL DGLWAFSLDRENCIDQRWWESALQESRAIAVPGSFNDQFADADI 60

40 Query: 76 RDYIGDFWYIEKVEVPKVSDEEMVLRFGSVTHQAKIYVDGVLVGEHKGGFTPFPEVLVPE 135  
 R+Y G+ WY++ + +PK + +VLR +VTH K++V+ V EH+GG+TPFE V  
 Sbjct: 61 RNYAGNVWYQREVFIPKGWAGQRIVLRFDVTHYGVWVNNQEVMEHQGGYTPFEADVTP 120

45 Query: 136 CKYNNEKIKVSI CANNVLDYTTLPVGNYS EIIQEDGSIKKKVRENDFDFNYAGVHRPLKL 195  
 + +++++C NN L++ T+P G I E+G KKK DFFNYAG+HR + L  
 Sbjct: 121 YVIAGKSVRITVCVNNELNWOQTIPGMV--ITDENG--KKKQSYFHDFDFNYAGIHRVML 176

50 Query: 196 MIRPKNHIFDITITSRLSDDLQSDADLHFLVETNQKVDEVRSVFDENKLV--GETKDSR 253  
 P + DIT+ + ++ D A + + V N +V + + D D ++V G+  
 Sbjct: 177 YTPNTWVDDITVVTHTVAQDCNHASVDWQVVAN---GDVSVELRDADQQVVATGQGTSGT 233

55 Query: 254 LFLSDVHLWEVLNAYLYTARVEIFVDNQLQDVYEENFGLREIEVTNGQFLNRPKPIYFKG 313  
 L + + HLW+ YLY V + D+Y G+R + V QFL+N KP YF G  
 Sbjct: 234 LQVVNPHLWQPGEGYLYELCVTAKSQTEC-DIYPLRVGIRSVAVKGEQFLINH KPFYFTG 292

Query: 314 FGKHEDTFINGRGLNEAANLMDLNLKDMGANSFR TSHYPYSEEMRLADRMGV LVIDEV 373  
 FG+HED + G+G + + D L+ +GANS+RTSHYPY+EEM+ AD G++VIDE  
 Sbjct: 293 FGRHEDADLRGKGF DNVLMVHDHALMDWIGANSYRTSHYPYAEEMLDWADEHGIVVIDET 352

Query: 374 PAVGLFQNFNASLDLS-----PKDNGTWNLM--QTKAAHEQAIQELVKRDKNHPSVVMW 425

-490-

AVG FN SL + PK+ + + +T+ AH QAI+EL+ RDKNHPSVVMW  
 Sbjct: 353 AAVG----FNLSLGIGFEAGNKPKELYSEEAVNGETQQAHLQAIKELIARDKNHPSVVMW 408

Query: 426 VVANEPASHEAGAHDYFEPLVKLYKLDLPQKRPVTLVNILMATPDRDQVMDLVDVCLNR 485  
 +ANEP + GA +YF PL + + LDP RP+T VN++ D + DL DV+CLNR  
 Sbjct: 409 SIANEPDTRPQGAREYFAPLAEATRKLDPTRPITCVNVMFCDAHTDTISDLFDVCLNR 467

Query: 486 YYGWYVDHGLDTNAEVGIRKELLEWQDKFPDKPIIITEYGADTLPLGHSTWNIPYTEEFQ 545  
 YYGWYV GDL AE + KELL WQ+K +PIIITEYG DTL GLHS + ++EE+Q  
 Sbjct: 468 YYGWYVQSGDLETAEKVLEKELLAWQEKL-HQPIIITEYGVDTLAGLHSMYTDWSEYQ 526

Query: 546 CDFYEMSHRVFDGIPNLVGEQVWNFADFETNLMILRVQGNHKGFLFSRNQPKQVVKEFKK 605  
 C + +M HRVPD + +VGEQVWNFADF T+ ILRV GN KG+F+R+R+PK +K  
 Sbjct: 527 CAWLDMYHRVFDVRSVAVVGEQVWNFADFATSQGILRVGNGKKGIFTRDRPKSAAFLLQK 586

Query: 606 RW 607  
 RW  
 Sbjct: 587 RW 588

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1259> which encodes the amino acid sequence <SEQ ID 1260>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -4.04 Transmembrane 1131 -1147 (1130 -1147)

----- Final Results -----

bacterial membrane --- Certainty=0.2614(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF97242 GB:AF282987 beta-galactosidase precursor [Streptococcus pneumoniae]  
 Identities = 303/921 (32%), Positives = 463/921 (49%), Gaps = 86/921 (9%)

Query: 5 QKSSEIVT---RTITKPSRATSNVQKEIDMTFDSKEQTVTGYYHYIDQ--EGRKQPFN 58  
 +K E VT + KP ++ + ++ ++Q E RK FN  
 Sbjct: 96 KKEDEAVTPKEEKVSAKPEEKAPRIESQASNQEKPLKEDAKAVTNEEVNQMIEDRKVDNF 155

Query: 59 QGWRF-LMADVACAQDPSFDDSNWQVIHLPHDFSLTQPYTRNGEA--ESAYKLGGVGWYR 115  
 Q W F L A+ A P D S W+ + LP+D+S+ + A E GG WYR  
 Sbjct: 156 QNWYFKLNANSKEAIKPDADVSTWKKLDLPYDWSIFNDFDHESPAQNEGGQLNGGEAWYR 215

Query: 116 HYLVLDEVLAGCHVAITFEFSYMETEIYVNGQFIGKHLNGYQEFTYDISDVVTF-GAENL 174  
 LDE +V +TF+G YM++++YVNGQ +G + NGY +F+YDI+ + G EN+  
 Sbjct: 216 KTFKLDEKDLKKNVRLTFDGVYVNGQVYVNGQLVGHYPNGYNQFSYDITKYLQKDGRENV 275

Query: 175 LAVRVENKVPSSRWYSGSGLYREVSLSVLPQLHFVADQVAMTLADTAVQEKQKQVDLRF 234  
 +AV NK PSSRWYSGSG+YR+V+L V ++H + + Q+ G+ + +  
 Sbjct: 276 IAVHAVNKQPSRWYSGSGIYRDVTLQVTDKVHVEKNGTTILTPKLEEQQHGKGVETHVTS 335

Query: 235 ALNQSIQTCHYQLSLCLWEQSHCSKDKKLLYQETEVPLADLAFQRQYGLT--LSLEELQL 292  
 + + H ++ E + + L L L +E +L  
 Sbjct: 336 KIVNTDDKDHELVA---EYQIVERGGHAVTGLVRTASRTLKAHESTSLDAILEVERPKL 391

Query: 293 WSP--DNPHLYDLELTLYYQGVVIDCFLETGFRQLTFMANQGLFVNGRAVKLGVCCLHH 350  
 W+ D P LY+L +Y GQ++D G+R + N+G +NG +K GV LHH  
 Sbjct: 392 WTVLNDKPALYELITRVYRDGQLVDAKKDLFGYRYHWTNPNEGFSLNGERIKFHGVSLLHH 451

Query: 351 DQGGLGACAYEDALARQLVLLKDMGANTIRSTHNPSSPKLRQLANRLGFFVIEEAFDTWT 410  
 D G LGA A R+L +K+MG N+IR+THNP+S + Q+A LG V EEAFDTW  
 Sbjct: 452 DHGALGAENYKAEYRRLKQMKEMGVNSIRTHNPASEQTLQIAELGLLVQEEAFDTWY 511

Query: 411 YAKNGNVNDFSNYFHQTIGTENANYLQVRSPETSWAQYSIEAMVWSAKNDPSVLMWSIG 470  
 K D+ +F + A ++ W+ + + MV KN+P++ MWSIG

-491-

Sbjct: 512 G GK--KPYDYGRFFFEKDATHPEARKGEK-----WSDFDLRTMVERGKNNPAIFMWSIG 562

Query: 471 NELMEGFSADVSHYPELTRQMCQWITAIDTSRPITFGDNKLKEADFC-WHEEVSQMATLL 529  
NE+ G + +H +++ + I +D +R +T G +K + + HE+++

5 Sbjct: 563 NEI--GEANGDAHSLATVKRLVKVIKDVKTRYVTMGADKFRFGNGSGGHEKIA----- 614

Query: 530 SQLDHPQGLIGLNYADGKDYDRLHEEHSWLLYGSETVSAITSR-AYYKETKKVLDSD--- 585  
+LD +G NY++ +Y L +H WL+YGSET SA +R +YY+ +++ S

10 Sbjct: 615 DELD----AVGFNYSE-DNYKALRAKHPKWLIIYGSETSSATRTGSIYRPERELKHSNGP 669

Query: 586 --GYHLTSYDHAKVDWGAFAASQAWYDTITRDFV--AGECVWTGFDYLGEPTPWNKIDSGV 641  
Y + Y + +V WG A+ +W T RD AG+ +WTG DY+GEPTPW+ +

15 Sbjct: 670 ERNYEQSDYGNDRVGWGTATASW--TFDRDNAGYAGQFIWTGTDYIGEPTPWHNQNTPT 727

Query: 642 VGLWSPSPKNAYFGILDTAGFPKDSYFYQSOW--AQQQTTLHLLPVWQKD-----QLCFD 694  
V K++YFGI+DTAG PK +Y YQSOW + + +HLLP W + D

Sbjct: 728 V-----KSSYFGIVDTAGIPKHDFLYQSQWVSVKKKPMVHLLPHWNWENKELASKVAD 781

20 Query: 695 EQGLVEVVVYSNAASVOLMFEDQGNLTDYGRKAFHTYSTPTGHTYQLYQGADAANKPHE 754  
+G + V YSNA+SV+L N G K F+ T G TYQ +GA+A

Sbjct: 782 SEGKIPVRAYSNAASSVELFL-----NGKSLGLKTFNKKQTS DGRTYQ--EGANA-----N 829

25 Query: 755 NLYLTWRVPYQKGLLRAVAYDISGKSIPKTSGRSQVRYTGSVAKLSWKAFAEAPIDAPW-E 813  
LYL W+V YQ G L A+A D SGK I R ++ T G A + + I A +

Sbjct: 830 ELYLEWKVAYQPGTLEAIARDESCKEI----ARDKITTAGKPAAVRLIKEDHAIADGKD 885

Query: 814 LLYLDLSLLDSRGELVSHAQDWLQVQVEGPARLLALDNGNPTDHTPYQEP-----LRQAY 868  
L Y+ ++DS+G +V A + ++ Q+ G +L+ +DNG Y+ +R+A+

30 Sbjct: 886 LTYIYIEIVDSQGNVPTANNLVRFLHGGQLVGVNDGEQASRERYKAQADGSGWIRKAF 945

Query: 869 GGKLLAILALTGEAGHIKVTA 889  
GK +AI+ T +AG +TA

35 Sbjct: 946 NGKGVAIVKSTEQAGKFTLTA 966

An alignment of the GAS and GBS proteins is shown below:

Identities = 98/414 (23%), Positives = 175/414 (41%), Gaps = 64/414 (15%)

40 Query: 54 LPSDEVMVIPSFNFLMVSKEKRDYIGDFWYEKVIEWPKVSEDEEMVLRFSGSVTHQAKIY 113  
LP D + P + N S K +G WY + + +V + + F + +IY

Sbjct: 86 LPHDFSLTQPYTRNGEABESAYKLGVG--WYRHYLVLDVLAGCHVAITFEQSYMETEIY 143

Query: 114 VDGVLVGEHKGGFTPFEVLVPECKYNNKIKVSICANNVLDYTTLPVGNYSIEIIQEDGSI 173  
V+G +G+H G+ F + + V+ A N+L +

45 Sbjct: 144 VNGQFIGKHLNGYQEFTYDISDV-----VTFGAENLLAVR-----V 179

Query: 174 KKKVRENDFDFNYAGVHRPLKLMIRPKNHIFDITITSRLSDDL-----QSADLHFLVET 227  
+ KV + +++ +G++R + L + P+ H + L+D Q DL F +

50 Sbjct: 180 ENKVPSS-RWYSGSGLYREVSLSVLPQLHFVADQVAMTLADTAVQEKGGQKVDLRFALNQ 238

Query: 228 NQKVDEVRIISVF-----DEDNKLVGSETKDS-----RLFLSDVHLWEVLNA 267  
+ + ++S+ +D KL+ + + L L ++ LW N

Sbjct: 239 SIQTCHYQLSLCLWEQSHCSKDKLLYQETEVPLADLAFQRYGLTSLLEELQLWSPDNP 298

55 Query: 268 YLYTARVEIFVDNQLQDVYEENFGLREIE-VTNGQFLNRPKIYFKGFGKHEDTFINGRG 326  
+LY + ++ Q+ D + G R++ + N +N + + KG H D G

Sbjct: 299 HLYDLELTLYYQGQVIDCFCLTGTFRQLTFMANQGLFVNGRAVKLGVCVCLHHDQGGGLGAC 358

60 Query: 327 LNEAANLMDLNLKDMGANSFRTSHYPYSEEMRLADRMGVLVIDEVPVAVGLFQ---NFN 383  
E A L LLKDMGAN+ R++H P S ++ +LA+R+G VI+E + N N

Sbjct: 359 AYEDALARQLVLLKDMGANTIRSTHNPSSPKLRQLANRLGFFVIEEAFDTWTYAKNGNVN 418

Query: 384 ASLDLSPKDNGTWN---LMQTKAAH---EQAIQELVKRDKNHPSVVMVWVANE 430  
+ + GT N L + ++ + +I+ +V KN PSV+MW + NE

65 Sbjct: 419 DFSNYFHQTIGTENANYLQVRSPETSWAQYSIEAMVWSAKNDPSVLMWSIGNE 472

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 387

A DNA sequence (GBSx0418) was identified in *S.agalactiae* <SEQ ID 1261> which encodes the amino acid sequence <SEQ ID 1262>. This protein is predicted to be 2-keto-3-deoxygluconate kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -0.53    Transmembrane  197 - 213 ( 197 - 213)

----- Final Results -----
          bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9699> which encodes amino acid sequence <SEQ ID 9700> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD35161 GB:AE001693 2-keto-3-deoxygluconate kinase [Thermotoga maritima]
Identities = 115/342 (33%), Positives = 180/342 (52%), Gaps = 16/342 (4%)

Query: 14  KIISLGEVLLRLSPPQYHTLMQANHLKQFGGSELNVLASLAQLGYHVGLVSALPDNDLG 73
          K+++ GE++LRLSPP + + Q + +GG+E NV A LAQ+G V+ LP+N LG
Sbjct: 2   KVVTFGEIMLRLSPPDHKRIHQDTSFDVTYGGAEANVAFLAQMGDAYFVTKLPNNPLG 61

Query: 74  KMASQFILSQQISPAAIKKEGRLGIYYEQGFSVRTNKVIYDRNYSSFWESTLSDYDFT 133
          A+ + + I + R+GIY+ E G S R +KV+YDR +S+ E+ D+D+
Sbjct: 62  DAAAGHLRKFVGKTDYIARGGNRIGIYFLEIGASQRPQKVVYDRAHSAISEAKREDFDWE 121

Query: 134 SIFKGVDFWVHSGITPALTKDLYEVTRFLMTKAKEGGVKVSIIDLNFRESLWSSFQEAREQ 193
          I G WFH SGITP L K+L + + A E GV VS DLN+R LW+ +EA++
Sbjct: 122 KILDGARWFHFGITPPLGKELPLILEDALKVANEGVTVSCDLNYRRLWTK-EEAQKV 180

Query: 194 LSPLLGLLDVCFGLEPIYLAGESEDLDKDELGLSRPYLDI-----ELLEKITQKIVQEY 246
          + P + +DV L ED++ LG+S LD+ E KI +++ ++Y
Sbjct: 181 MIPFMEYVDV-----LIANEEDIEKVLGISVEGLDLKTGKLNREAYAKIAEEVTRKY 232

Query: 247 GLDYIAFTQREMEYTNQYMLKSYLYHNMLYQTDKTGVEVLDRVGTGDAFAAGLIHALLE 306
          + T RE ++ N + +++ + ++DRVG GD+FA LI+ L
Sbjct: 233 NFKTVGITLRESISATVNYWSVMVFENGQPHFSNRYEIHIVDRVAGDSFAGALIYGSLM 292

Query: 307 KETPQRALEIAMATFKYKHTIQGDINIMTRDDIAYLIEKETN 348
          Q+ E A A KHTI GD +++ ++I L T+
Sbjct: 293 GFDSQKKAEEFAAASCLKHTIPGDFVLSIEEIEKLAGATS 334
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1263> which encodes the amino acid sequence <SEQ ID 1264>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.0708(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 111/319 (34%), Positives = 168/319 (51%), Gaps = 7/319 (2%)
```

-493-

Query: 12 MAKIISLGEVLLRLSPFQYHTLMQANHLKCQFGGSEINVLASLAQLGYHVGLVSALPDND 71  
 M+K++ +GE L+R+SP Q+ L A + FGGSE+N+ +L G L +ALPDN  
 Sbjet: 14 MSKLLLVGEPLIRVSPNQFQPLTNACEAQLFFGGSEVNIARTLGGFGLEARLFTALPDNP 73

Query: 72 LGKMASQFILSQQISPAAIKKEGRIGIYYEQQGFSVRTNKVIYDRNYSSFWESTLSDDYD 131  
 +G QF+ + + + R+G+YY E GF R ++V YDR SSF D  
 Sbjet: 74 VGHAFHQFLKQSGVMSLTAWQGHVRGLYYLENGFGCRASQVYDRCGSSFSALDKDLSLD 133

Query: 132 FTSIFKGVDFHVSGITPALTKDLYEVTRFLMTKAKEGGVKVSIDLNFRESLWSSFQEAR 191  
 +IF+G+ FH SGI+ AL K ++ L+ +AK+ + +S DLNFR S+ + +A+  
 Sbjet: 134 LAAIFEGISHFHFSGISLALGKKTQDLIEVLVREAKKRDICISFDLNRSSM-IAVADAK 192

Query: 192 EQLSPLGLLDVCFGLEPIYLAGESEDLKDELGLSRPYLDIELLEKITQKIVQEYGLDYI 251  
 S D+ FG+EP+ L + D+ D R D + + + Q Y L I  
 Sbjet: 193 RLFShFAQYADIIFGMEPLLLDSDDFDMFD-----RKKADTTTIRERLAGLYQRYQLQAI 247

Query: 252 AFTQREMEYTNQYMLKSYLYHNNMLYQTDKTGVEVLDRVGTGDAFAAGLIHALLEKETPQ 311  
 T+R + K+Y Y + Y++ + VL RVG+GDAF AGL++ LLE Q  
 Sbjet: 248 YHTERSNDAGQSNHFKAYAY-DRQFYESCEVTTPVLQRVGSGDAFVAGLLYQLLEGNEKQ 306

Query: 312 RALEIAMATFKYKHTIQGD 330  
 R L+ A+AT K T+ D  
 Sbjet: 307 RNLDFAVATASLKCTVAED 325

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 388

A DNA sequence (GBSx0419) was identified in *S.agalactiae* <SEQ ID 1265> which encodes the amino acid sequence <SEQ ID 1266>. Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -1.17 Transmembrane 5 - 21 ( 5 - 21)

----- Final Results -----  
 bacterial membrane --- Certainty=0.1468(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 389

A DNA sequence (GBSx0420) was identified in *S.agalactiae* <SEQ ID 1267> which encodes the amino acid sequence <SEQ ID 1268>. Analysis of this protein sequence reveals the following:

Possible site: 53  
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -12.05	Transmembrane	198 - 214 ( 191 - 220)
INTEGRAL	Likelihood = -11.68	Transmembrane	446 - 462 ( 437 - 467)
INTEGRAL	Likelihood = -9.55	Transmembrane	94 - 110 ( 91 - 116)
INTEGRAL	Likelihood = -7.43	Transmembrane	291 - 307 ( 283 - 309)
INTEGRAL	Likelihood = -4.88	Transmembrane	265 - 281 ( 257 - 282)
INTEGRAL	Likelihood = -4.62	Transmembrane	321 - 337 ( 318 - 339)
INTEGRAL	Likelihood = -3.93	Transmembrane	406 - 422 ( 405 - 426)
INTEGRAL	Likelihood = -1.59	Transmembrane	121 - 137 ( 121 - 137)

-494-

INTEGRAL Likelihood = -1.12 Transmembrane 345 - 361 ( 345 - 362)  
 INTEGRAL Likelihood = -0.48 Transmembrane 43 - 59 ( 43 - 59)

## ----- Final Results -----

5 bacterial membrane --- Certainty=0.5819(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:CAB13641 GB:Z99113 similar to H+-symporter [Bacillus subtilis]  
 Identities = 105/452 (23%), Positives = 182/452 (40%), Gaps = 37/452 (8%)

Query: 36 IYLFTFMFVTYFSTGVLGSAIFVSQIMGYIRIFDGFIDPAIGIMIDKTDTKFGKYRPIL 95  
 IY ++ +F T V G +A + +RI D DP IG ++D+T+++F ++RP L

15 Sbjct: 27 IYATVSTYLLFFYTDFVGLSAAAAGTMFLVVRIDALADPFIGTIVDRTNRSRFRFRPYL 86

Query: 96 IIGNVITALSLIFLLALRGVDENIRFPLFILVLIHKIGYSMQQTITKAGQTALTNDPKQ 155  
 + G A + L L + ++ I +G S+ T ALT+

20 Sbjct: 87 LFG----AAPPVILAILCFTTDFSDMGKLIYAYTTYVGLSLTYTTINVPGALTS-AMT 141

Query: 156 RPIFNIVDAVMTTSLMTGGQFVSVFLVPKFGNFTPQFFNVLI FG TILISAILAIV--AI 213  
 R +V L +V F VP + G L IL ++ +

25 Sbjct: 142 RNNQEVVSITSVRMLFANLGLVVAFFVPLLAAYLSDTSGNESLQWQLTMGILGMIGGCL 201

Query: 214 IGIWAKDRKEFFGLGENTQKTALKDYWKVLKGNKPLQILSIAAALVKFAIQFFGDSV-VM 272  
 + K KE L ++ +K D ++ + N+PL +LSI ++ F + +SV +

30 Sbjct: 202 LIFCFKSTKERVTLQKSEEKIKFTDIFEQFRVNRPLVVLISIFFIII-FGVNSISNSVGIY 260

Query: 273 VLLFGI----LFGNYALSGQFSLLFIVPGVIINILFSTIARKKGLRFSYVRAIQIGMIGL 328  
 + + + L Y L G L I+P I L + +KK L + A+ + +IGL

35 Sbjct: 261 YVTYNLEREDLVKWYGLIGSLPALVILP--FIPRLHQFLGKKKLLNY----ALLLNIIGL 314

Query: 329 LAFGAVLYVGKPGDLSLTSNLTYTILFIVTNI IARYASQAPASLVLTMGADISDYETSES 388  
 LA L + N+Y IL V +IA S + + + +Y +

35 Sbjct: 315 LAL-----LFVPPSNVYLIL--VCRLIAAAGSLTAGGYMWALIPETIEYGEYRT 361

Query: 389 GRYVSGMIGTIFSLTDSIASSFAPMVVGVLAGIGFSKSFPTIETPLPPDLKMAAISILV 448  
 G+ + G+I I + +V G VL G+ P M +

40 Sbjct: 362 GKRMGGLIYAIIGFFFKFGMALGGVVPGLVLDKFGY-----VANQAQTPAALMGILITTT 416

Query: 449 AIPFIALSIALLLMKFYKLDKEEMVRIQEKIQ 480  
 IP L +AL+ + FY LD+++ + +++

Sbjct: 417 IIPVFLVLALIDINFYNLDEKKYKNMVRELE 448

45 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 390**

50 A DNA sequence (GBSx0422) was identified in *S.agalactiae* <SEQ ID 1269> which encodes the amino acid sequence <SEQ ID 1270>. Analysis of this protein sequence reveals the following:

Possible site: 52  
 >>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.3375(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:



-495-

>GP:AAB17663 GB:U31175 D-specific D-2-hydroxyacid dehydrogenase [*S. aureus*]  
Identities = 165/331 (49%), Positives = 231/331 (68%), Gaps = 1/331 (0%)

Query: 1 MMKLKVFNVREEEATLAQDWANRNHVELSMSEGPLTLETVNEVEGFDGIANAQIEPLDDA 60  
M K+ F R+ E +A +W +N+VE++ S+ L+ TV++++ +DG+ Q L++  
Sbjct: 1 MTKIMFFGTRDYEKEMALNWGKKNVETTSKELLSSATVDQLKDYDGVTTMQFGKLEND 60

Query: 61 IYPLLKEMGIKQIAQRSAGVDMYNLELAKQHGIISNVPSYSPEIAEFTVTIALNLIRK 120  
+YP L+ GIKQIAQR+AG DMY+L+LAK+H I+ISNVPSYSPE+IAE++V+IAL L+R+  
Sbjct: 61 VYPKLESYGIKQIAQRTAGFDMYDLDLAKKHIVISNVPSYSPETIAEYSVSIALQLVRR 120

Query: 121 VELIRANVREQNFSWTLPIRGRVLGNMTVAIIIGTGRIGLATAKIFKGFGRVIGYDIYHN 180  
I V+ +F+W I + + NMTVAIIIGTGRIG ATAKI+ GFG + YD Y N  
Sbjct: 121 FPDIERRVQAHDFTWQAEIMSKPVKNMTVAIIIGTGRIGAATAKIYAGFGATITAYDAYPN 180

Query: 181 PMADGILEYVNSVEEAVEEADLVSLHMPPTAENTHLFNLDMPKQFKKGAILNMARGALV 240  
D L Y +SV+EA+++AD++SLH+P E+ HLF+ MF KKGAIL+N ARG++  
Sbjct: 181 KDLD-FLTYKDSVKEAIKDADIISLHVPANKESYHLFDKAMFDHVKKGAILVNAARGAVI 239

Query: 241 ETKDLLEALDQGLLEGAGIDTYEFGPYIPKNCQQDISDKDFLRRLINHPKVIYTPHAAY 300  
T DL+ A++ G L GA IDTYE E Y + +DI DK L LI H +++ TPH A+  
Sbjct: 240 NTPDLIAAVNDGTLGAAIDTYENEAAYFTNDWINKDIDDKTLELIEHERILVTPHIAF 299

Query: 301 YTDEAVKNLVEGALNACVEVIETGTTTTTKVN 331  
++DEAV+NLVEG LNA + VI TGT T++N  
Sbjct: 300 FSDEAVQNLVEGGLNAALSVINTGTCTETRLN 330

There is also homology to SEQ ID 124.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 391

A DNA sequence (GBSx0423) was identified in *S.agalactiae* <SEQ ID 1271> which encodes the amino acid sequence <SEQ ID 1272>. Analysis of this protein sequence reveals the following:

Possible site: 40  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2364(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 392

A DNA sequence (GBSx0424) was identified in *S.agalactiae* <SEQ ID 1273> which encodes the amino acid sequence <SEQ ID 1274>. This protein is predicted to be regulatory protein (pfoS/R). Analysis of this protein sequence reveals the following:

Possible site: 37  
>>> Seems to have a cleavable N-term signal seq.  
INTEGRAL Likelihood =-12.90 Transmembrane 64 - 80 ( 53 - 89)

----- Final Results -----

-496-

bacterial membrane --- Certainty=0.6158(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 9325> which encodes amino acid sequence <SEQ ID 9326> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema  
 pallidum]  
 Identities = 33/91 (36%), Positives = 55/91 (60%), Gaps = 1/91 (1%)  
 Query: 1 MANVLAKPKIMLPMISSAAILGILGALFNIQGTPASAGFGISGLIGPINALNLAKGWSV 60  
 M N + P + +P++ + + G+L LFN+QGTPASAGFG GL+GPINA L V  
 15 Sbjct: 250 MPNWIRYPILNIPLLLNGLVCGVLAWLFNLQGTPASAGFGFIGLVGPINAYRLMAYTPMV 309  
 Query: 61 MNMLLIIFVAAPIILNFIFNYLFIKVLKI 91  
 +L ++ FV + + ++ +++ + LK+  
 Sbjct: 310 RAGILFLVYFVLS-FLAAYLIDFILVDRLKL 339

- 20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1275> which encodes the amino acid sequence <SEQ ID 1276>. Analysis of this protein sequence reveals the following:

Possible site: 51  
 >>> Seems to have a cleavable N-term signal seq.  
 25 INTEGRAL Likelihood = -12.31 Transmembrane 141 - 157 ( 133 - 166)  
 INTEGRAL Likelihood = -6.00 Transmembrane 92 - 108 ( 88 - 112)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5925(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

35 >GP:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema  
 pallidum]  
 Identities = 63/178 (35%), Positives = 107/178 (59%), Gaps = 10/178 (5%)  
 Query: 2 IGQGIASLLGLQPILMSLLIAMIFCFLIVSPITTVGIALAINLSGIGSGAASFG----- 55  
 +G+ IA+ + LQP+LMS+L++M F +I+SP+++V + +A+ L+G+ SGAA+ G  
 40 Sbjct: 164 VGRVIATFIALQPLLSILLSMSFSLLIISPVSVAVGIAVGLTGLASGAANIGVSSCAM 223  
 Query: 56 -LCLAGWAVNSKGTSLAHVLRSPKISMANVLSKPKIMLPMLCSAAVLGVIGAI FNQGTP 114  
 L + VN G LA + K+ M N + P + +P+L + V GV+ +FN+QGTP  
 Sbjct: 224 TLIVGTM RVNKGIVPLAMFAGAMKMLMPNWIRYPILNIPLLLNGLVCGVLAWLFNLQGTP 283  
 45 Query: 115 ASAGFGISGLIGPINALNLAKGGWCP-VNILLIIIFVGAPIVLNMIFNYLFIKVLKV 171  
 ASAGFG GL+GPINA L + P V ++ +++ + + +++ + LK+  
 Sbjct: 284 ASAGFGFIGLVGPINAYRLM--AYTPMVRAGILFLVYFVLSFLAAYLIDFILVDRLKL 339

An alignment of the GAS and GBS proteins is shown below:

50 Identities = 86/101 (85%), Positives = 96/101 (94%)  
 Query: 1 MANVLAKPKIMLPMISSAAILGILGALFNIQGTPASAGFGISGLIGPINALNLAKGWSV 60  
 MANVL+KPKIMLP+ SAA+LG++GA+FNIQGTPASAGFGISGLIGPINALNLAKGGW  
 55 Sbjct: 81 MANVLSKPKIMLPMLCSAAVLGVIGAI FNQGTPASAGFGISGLIGPINALNLAKGGWCP 140  
 Query: 61 MNMLLIIFVAAPIILNFIFNYLFIKVLKIIDPMDYKLDI 101  
 +N+LLIIIFV API+LN IFNYLFIKVLK+IDPMDYKLDI  
 Sbjct: 141 VNILLIIIFVGAPIVLNMIFNYLFIKVLKVIDPMDYKLDI 181

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 393

A DNA sequence (GBSx0426) was identified in *S.agalactiae* <SEQ ID 1277> which encodes the amino acid sequence <SEQ ID 1278>. This protein is predicted to be regulatory protein (pfoS/R). Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -6.58	Transmembrane	148 - 164 ( 145 - 169)
INTEGRAL	Likelihood = -5.26	Transmembrane	33 - 49 ( 25 - 52)
INTEGRAL	Likelihood = -4.73	Transmembrane	70 - 86 ( 62 - 88)
INTEGRAL	Likelihood = -3.45	Transmembrane	124 - 140 ( 122 - 143)
INTEGRAL	Likelihood = -1.33	Transmembrane	96 - 112 ( 96 - 112)

----- Final Results -----

bacterial membrane	---	Certainty=0.3633(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 9735> which encodes amino acid sequence <SEQ ID 9736> was also identified.

A related GBS nucleic acid sequence <SEQ ID 9697> which encodes amino acid sequence <SEQ ID 9698> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema pallidum]  
Identities = 61/158 (38%), Positives = 92/158 (57%)

Query: 24	KSFIMNVNLGLALGTVIVLIPGAILGELMKALLPMWSGFATLIAATAVATSMGLVIGIM 83
	+ F+M +LNG + G VI L+P AI GEL +AL P+ FA L + +IG +
Sbjct: 9	RQFMMKILNGSSAGIVIGLVPPAIALGELFRALAPLSPLFAALYHVVLPIQFSVPALIGTL 68

Query: 84	VGLNFKFNPIQSASLGLAVMFAGGAATFLKGAIMLKGTGDIINMGITAALGVLLIQFLSD 143
	VGL F + + A+L + A G T GA ++ G GD+IN+ + +AL ++L++ L

Sbjct: 69	VGLQFHCSAPEVATLAFVSVIASGNVTLQNGAWLITGIGDVINVMLISALAILVRALRG 128
-----------	---

Query: 144	KTKSFTLIVIPTVTLLLVGGVGHVLLPYVKMITTMIGQ 181
	K S T+I +P + ++ GGVG LPYVKMIT +G+
Sbjct: 129	KLGSLLTIALPVIVAVVAGGVGSFSLPYVKMITLTVGR 166

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1279> which encodes the amino acid sequence <SEQ ID 1280>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -13.06	Transmembrane	314 - 330 ( 301 - 335)
INTEGRAL	Likelihood = -11.30	Transmembrane	185 - 201 ( 178 - 215)
INTEGRAL	Likelihood = -8.01	Transmembrane	22 - 38 ( 11 - 42)
INTEGRAL	Likelihood = -3.29	Transmembrane	266 - 282 ( 265 - 285)
INTEGRAL	Likelihood = -2.66	Transmembrane	141 - 157 ( 141 - 159)
INTEGRAL	Likelihood = -2.13	Transmembrane	53 - 69 ( 53 - 69)
INTEGRAL	Likelihood = -1.33	Transmembrane	114 - 130 ( 113 - 131)
INTEGRAL	Likelihood = -0.80	Transmembrane	206 - 222 ( 206 - 222)

----- Final Results -----

bacterial membrane	---	Certainty=0.6222(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

-498-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the databases:

5 >GP:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema  
pallidum]  
Identities = 137/346 (39%), Positives = 217/346 (62%), Gaps = 14/346 (4%)

10 Query: 12 FMNKVLGTAIAIVVALIPNAILATFLKPLLP-NMAAAEFLHIVQVFQFFTPIMAGFLIG 70  
FM K+L G++ IV+ L+P AI + L P + A H+V QF P + G L+G  
Sbjct: 11 FMMKILNGSSAGIVIGLVPPAIGELFRALAPLSPLFAALYHVVLPIQFSVPALIGTLVG 70

15 Query: 71 QQKFKNPMQQLAVGGAAYIGSGAWAYTEVIQKGVATGTFQLRGIGDLINMMITASLAVLA 130  
QF + + + + I SG + G + + GIGD+IN+M+ ++LA++  
Sbjct: 71 LQFHCSAPEVATLAFVSVIASG-----NVTLQNGAWLITIGDVINVMLISALAIL 122

20 Query: 191 IAVAFSLIIVSPISTVAIGLAIGLNGMAAGAASMGIASTA AVLWVATLKVNKSGVPIAIA 250  
++++FSLII+SP+S+VA+G+A+GL G+A+GAA++G++S A L+ T++VNK GVP+A+  
Sbjct: 183 LSMSFSLIISPVSSVAVGIAVGLTGLASGAANIGVSSCAMTLIVGTMRVNKIGVPLAMF 242

25 Query: 251 LGAMKMMMPNFKHPIMAIPMVFTAAISSLTVPFLNLTGTPASSGFLVGAVGPIAS--L 308  
GAMKM+MPN++++PI+ IP++ + + LFNLTGPAS+GFG +G VGPI + L  
Sbjct: 243 AGAMKMLMPNWIRYPILNIPLLNLGVLAWLFLNLQGTASAGFGFIGLVGPINAYRL 302

30 Query: 309 AGGSSIL---IILAWIIVPFAVAFAAHKVSKDILKLYKEDIFVFE 351  
+ ++ I+ L + ++ F A+ + D LKLY+ ++F+ E  
Sbjct: 303 MAYTPMVRAGILFLVYFVLSFLAAYLIDFILVDRLLKLYRRELFIE 348

An alignment of the GAS and GBS proteins is shown below:

Identities = 65/172 (37%), Positives = 95/172 (54%), Gaps = 9/172 (5%)

35 Query: 19 EKQTTKSFIMNVNLGLALGTIVIVLIPGAILGELMKALLPMWSGFATLIAATAVATSMGL 78  
+K+T SF+ VL G A+ V+ LIP AIL +K LLP + A + V +  
Sbjct: 5 DKETFFSFMNKVLGTAIAIVVALIPNAILATFLKPLLPNMAA-AEFLHIVQVFQFFTP 63

40 Query: 79 VIGIMVGLNFKFNPIQSASLGLAVMFAGGAATFLK-----GAIMLKGTGDIINMGIT 130  
+ G ++G PKFNP+Q ++G A GA + + G L+G GD+INM IT  
Sbjct: 64 MAGFLIGQQKFKNPMQQLAVGGAAYIGSGAWAYTEVIQKGVATGTFQLRGIGDLINMMIT 123

45 Query: 131 AALGVLLIQFLSDKTKSFTLIVIPTVTLILLVGGVGHVLLPYVKMITTMIGQG 182  
A+L VL +++ +K S T+I++P VG +G LPYV +TT+IGQG  
Sbjct: 124 ASLAVLAVKYFGNKFGLSLTIILLPITIGTGVGYIGWKFLPYVSYVTTLIGQG 175

A related GBS gene &lt;SEQ ID 8567&gt; and protein &lt;SEQ ID 8568&gt; were also identified. Analysis of this protein sequence reveals the following:

50 Lipop: Possible site: -1 Crend: 10  
McG: Discrim Score: -13.49  
GvH: Signal Score (-7.5): -5.82  
Possible site: 48  
>>> Seems to have no N-terminal signal sequence

55 ALOM program count: 5 value: -6.58 threshold: 0.0

INTEGRAL	Likelihood = -6.58	Transmembrane	148 - 164 ( 145 - 169)
INTEGRAL	Likelihood = -5.26	Transmembrane	33 - 49 ( 25 - 52)
INTEGRAL	Likelihood = -4.73	Transmembrane	70 - 86 ( 62 - 88)
INTEGRAL	Likelihood = -3.45	Transmembrane	124 - 140 ( 122 - 143)
INTEGRAL	Likelihood = -1.33	Transmembrane	96 - 112 ( 96 - 112)
PERIPHERAL	Likelihood = 1.85		51

60 modified ALOM score: 1.82

\*\*\* Reasoning Step: 3

-499-

----- Final Results -----

bacterial membrane --- Certainty=0.3633(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF01226(352 - 843 of 1218)

EGAD|138195|TP0038(3 - 166 of 350) regulatory protein {Treponema pallidum} OMNI|TP0038  
 regulatory protein (pfoS/R) GP|3322295|gb|AAC65034.1|AE001189 regulatory protein (pfoS/R)  
 {Treponema pallidum} PIR|E71373|E71373 probable regulatory protein (pfoS/R) - syphilis  
 spirochete

%Match = 13.6

%Identity = 37.2 %Similarity = 59.1

Matches = 61 Mismatches = 67 Conservative Sub.s = 36

```

273      303      333      363      393      423      453      483
I*FFPIFLLQIAMI*LI*LVKSQTIIISRRHMSDVVEKQTKSFIMNVNLGLALGTIVIVLIPGAILGELMKALLPMWSG
      : : : |:| :||| : | || |:| || ||| :|| |:|
MHTQSLSPRQFMKILNGSSAGIVIGLVPPAIAGELFRALAPLSPL
                        10      20      30      40

513      543      573      603      633      663      693      723
FATLIAATAVATSMGLVIGIMVGLNFKFNPIQSASLGLAVMFAGGAATFLKGAIMLKGTGDIINMGITAALGVLLIQFL
||| : :|| :||| | : : |:| : : | | |: || : : | |:| |:| : : || : : || : : ||
FAALYHVVLPPIQFSVPALIGTLVGLQPHCSAPEVATLAFVSVIASGNVTLQNGAWLITGIGDVINVMLISALAILVRAL
      60      70      80      90      100      110      120

753      783      813      843      873      903      933      963
SDKTKSFTLIVIPVTVLLLVGGVGHVLLPYVKMITTMIGQGTRRTHENFLFILLCPDINFEEKIPF*INDLLSLFLQIIGL
| |:| :| : : : |||| ||||| :| :|
RGKLGSLTIIALPVIVAVVAGGVGSFSLPYVKMITLTVGRVIATFIALQPLLMSILLSMSFSLIIISPVSVAVGIAVGL
      140      150      160      170      180      190      200

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 394

A DNA sequence (GBSx0428) was identified in *S.agalactiae* <SEQ ID 1281> which encodes the amino acid sequence <SEQ ID 1282>. This protein is predicted to be cyn operon transcriptional activator. Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB15857 GB:Z99123 alternate gene name: ipa-24d-similar to  
 transcriptional regulator (LysR family) [Bacillus subtilis]  
 Identities = 87/282 (30%), Positives = 152/282 (53%), Gaps = 5/282 (1%)

Query: 1 MDIRQLTYFIAVAEAKNYSRAAKSLFVTQPTLSQSIKKLEAELNTILFLQNGRQLALTEA 60

MDIR LTYF+ VA K++++A++SL+V+QPT+S+ IK LE EL LF +NGRQ+ LT+A

Sbjct: 1 MDIRHLTYFLEVARLKSFTKASQSLYVSQPTISKMIKNLEELGIELFYRNGRQVELTDA 60

Query: 61 GEILYEKGQQLMTNVNQMVTEIQQLNQEKKEGIRVGLTSLFAIQFMKQI-STFMATHSNV 119

G +Y + Q ++ + + +E+ + + KK +R+GL + F ++ F + NV

Sbjct: 61 GHSMYVQAQEIIKSFQNLTSLENDIMEVKKGHVRIGLPPMIGSGFFPRVLGDFRENYPNV 120

-500-

Query: 120 EVSLIQDGSRKLOELLAKGKIDIGLLSFPSTRNDITIEPLQTSTKGYKVSIVMPKSHPLA 179  
 L++DGS K+QE + G +DIG++ P+ + + T + +V+ SH LA  
 Sbjct: 121 TFQLVEDGSIKVGEGVGDGSLDIGVVLPANEDIFHSFTIVKET----LMLVVHPSHRLA 176

5 Query: 180 TLPEIELNDLRDYKVASLNEHYMLGEMLPKRCRALGFDPHIVFKHNDWEVLIHSLQDLNA 239  
 E +L +L+D E ++L + +C GF PHI+++ + W+ + +  
 Sbjct: 177 DEKECQLRELKDEPFIFFREDFVLHNRIMTECIKAGFRPHIYETSQWDFISEMVSANLG 236

10 Query: 240 VTILPSEFESISQVQDLCWVPLKDKNNFYPIGIAYRNDTSFS 281  
 + +LP + + +PL D + + I +R D S  
 Sbjct: 237 IGLLPERICRGLDPEKVKVIPLVDPVIPWHLAIWRKDRYLS 278

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1283> which encodes the amino acid sequence <SEQ ID 1284>. Analysis of this protein sequence reveals the following:

15 Possible site: 21  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1101(Affirmative) < succ>  
 20 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 125/160 (78%), Positives = 144/160 (89%)

25 Query: 135 LAKGKIDIGLLSFPSTRNDITIEPLQTSTKGYKVSIVMPKSHPLATLPEIELNDLRDYKV 194  
 L++GKIDIGLLSF S R DITIE LQTSTKGYKVSIV+ K HPLA P+++L DL+ YK+  
 Sbjct: 1 LSQGKIDIGLLSFLSIRKDITIELLQTSTKGYKVSIVLLKQHPLAQHPQLKLDLKGKYI 60

30 Query: 195 ASLNEHYMLGEMLPKRCRALGFDPHIVFKHNDWEVLIHSLQDLNAVITILPSEFESISQVQ 254  
 ASLN+HYMLGEMLPKRCRALGF+P IVFKHNDWEVLIHSL DLN +TILPS+FES++QV  
 Sbjct: 61 ASLNDHYMLGEMLPKRCRALGFEPDIVFKHNDWEVLIHSLHDLNTLTILPSDFESLNQVD 120

35 Query: 255 DLCWVPLKDKNNFYPIGIAYRNDTSFSPMIEEFLSLLKTN 294  
 +L W+PL+DKNNFYPIGIAYR+D SFSP+IEEFLSLLKTN  
 Sbjct: 121 NLVWIPLQDKNNFYPIGIAYRDDASFSPVIEEFLSLLKTN 160

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### 40 Example 395

A DNA sequence (GBSx0429) was identified in *S.agalactiae* <SEQ ID 1285> which encodes the amino acid sequence <SEQ ID 1286>. Analysis of this protein sequence reveals the following:

Possible site: 47  
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1833(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 Signal peptide: 1-21

A related GBS nucleic acid sequence <SEQ ID 8569> which encodes amino acid sequence <SEQ ID 8570> was also identified.

55 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.